

RESULT 6
 TVCHLV
 epidermal growth factor receptor precursor - chicken
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erdb
 C:Species: Gallus gallus (chicken)
 C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C:Accession: A27720; A00643
 R:Latx, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr
 Mol. Cell. Biol. 8, 1970-1978, 1988
 A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mol
 A:Reference number: A27720; MUID:88261272; PMID:3260329
 A:Accession: A27720
 A:Molecule type: mRNA
 A:Residues: 1-1223 <LAX>
 R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
 Cell 41, 719-726, 1985
 A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:85228222; PMID:2988784
 A:Accession: A00643
 A:Molecule type: mRNA
 A:Residues: 585-1223 <NIL>
 A:Cross-references: GB:M1066
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor: protein kinase homology
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 C:Specific: protein kinase
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:31-654/Domain: extracellular #status predicted <EXT>
 F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
 F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:777-795/Region: protein kinase ATP-binding motif
 F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
 F:667/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query March 44.6%; Score 3053.5; DB 1; Length 1223;
 Best Local Similarity 48.2%; Pred. No. 1.7e-119;
 Matches 631; Conservative 169; Mismatches 347; Indels 161; Gaps 29;

QY 8 RWGLLALLPPGAA-----STQVCTGTMKLRLPASPTHLDMLRHLVYGGCQVYQGNLE 61
 DB 13 RGAADVLLVLLGLVALCSAVEEKVCCGGINNKLTOLGHEHFTSLQGMVNNCEVLSNLE 72
 QY 62 LTYPTNASLSFLDIOEVOGYVLIANNOVROVLOLRIRVGTQGFEDVYALAVDNGD 121
 DB 73 ITVEHNRDLFTLTIOEVAGYVLIANMDVPLENLIIRGNVLTIDNSFALAVSNYH 132
 QY 122 PLNNTPTVTSAPGSLRELOLRSLTEILKGVLIQRNPOLCYDPTILMKDIFKNNQAL 181
 DB 133 -MKTKQ-----GLRELPMKRSLSEILNGSVKISNNPKLCNMDTVLMNDIIDSRR-PL 182
 QY 182 TLID-TNRSAACHPCSPMGKSGRWGSSSEDCCSLTTCVAGCA-RCKRPLTDCHEQ 229
 DB 183 TVLDFASNLSGCPKCHNCTEDHGMAGBQNCQTLTVICAOQCSGCRKQVSDCHNQ 242
 QY 240 CAAGCTGPKHSDIACLHFNHSGICEALCPALVYNTDTESPMPREGRYTFASCVTAC 299
 DB 243 CAAGCTGPRSDCLAGKFRDADATCTCPPLVLYNTTYQMDNPNRGKTSFATCTREC 302
 QY 300 PNYLTSDVSGCTIVCGLAHNOEVAEDGTORCEKSGPCARVCTGLAQYIKANSKFIGI 359
 DB 303 PHNVVTVDHGSCVRSCTDITYEV-EENGVRKCKKCDLGSKVCNGTIGELKGLIS-INA 360
 QY 360 TELE-FAGCKKIFGSLAFLESFDDPASNTAPLPQPLQVFTLEITGYLISAMPDS 418

DB 361 TNIDSPKNTKINGDVSIIPVAFLGDAFTKLPDPKLDVFRVVEISGFLLIQMPON 420
 QY 419 LPDLVSFQNLQVIRGRIILHNCAVSLTLQGLISWIGLSRLRELSGSLALIHNTHLCPYH 478
 DB 421 AIDLVAFEULLEIRGTRKHQGYSLAVVNLKIQSLGRBLKISDSDIMMKNKLCYAD 480
 QY 479 TVPMDLPFNPHQALLHTANRPDECEGGLACHQICARGHCGWPGPFTCCVNSQFLRQ 538
 DB 481 TMMVRSLEFATQSQTKIIONRNKNDCTADRHVCDPLCSDVCGWGPFPFHCSCRFPSRQ 540
 QY 539 ECVEBRVIOGLPREVYNAHCLPCPREQOPONG---STCGPREADOCVCAHYDPPF 595
 DB 541 ECVQGNITQGEERERDSKCLPCSEBCLVNSTAVNTYTCGPGPHCMKCAHFTDGFH 600
 QY 596 CVARCPF---NNFTVSWLRVPKVSASHLEBDEGACOPCEINCTHSCVDLDDKCPAE 651
 DB 601 CVACGAGVLGENDTL-VWKYA-----DANAVCQLHPNCTRGCKGGLGECF-- 647
 QY 652 QRASPLTSIVAVV-GILLVVLGVVFGILIRROQKIRKYMRLLOTELVEPLTPSG 710
 DB 648 -NGSKTPSIAGVVGGLCLVVGGLIGLYLRR-HVAKRTLRRLORERLEVEPLTPSG 705
 QY 711 AMPNQAKRILKETELRKVYLGSGAFVYVYGMIPDGENYKIPVAIKVLRBENTSPKAN 770
 DB 706 EAPNOAHLRIKETEFKTKVYLGSGAFVYVYGMIPBEEKYKIPVAIKVLRBENTSPKAN 765
 QY 771 KEILDEAYMAGVSPYRLLGICLTSTVQVLTQDLMPGCLLDHVENRGSLGSDLLN 830
 DB 766 KEILDEAYMAGVSPYRLLGICLTSTVQVLTQDLMPGCLLDHVENRGSLGSDLLN 825
 QY 831 WCMQIAKMSYLEDVLYNRDLAARVLYKSNHYKITPGLARLLIDETVYHMDGVV 890
 DB 826 WCMQIAKMSYLEDVLYNRDLAARVLYKSNHYKITPGLARLLIDETVYHMDGVV 885
 QY 891 PIKMALESILRRFTQSDVSYGVYVWELMTFGAKPYDGIPIREIPDLLEKGERLP 950
 DB 886 PIKMALESILRRFTQSDVSYGVYVWELMTFGAKPYDGIPIREIPDLLEKGERLP 945
 QY 951 PICTIDVYIMKCMWIDSECPRRRELVESESRKARDQRFVYIQ-NEDLGPASPLDST 1009
 DB 946 PICTIDVYIMKCMWIDSECPRRRELVESESRKARDQRFVYIQ-NEDLGPASPLDST 1005
 QY 1010 FRSILLEDMDGDLVDAEFLYLPVQGFPCPDPAAGGVHHRHSSSTRSGGDLTGL 1069
 DB 1006 FRSILLEDMDGDLVDAEFLYLPVQGFPCPDPAAGGVHHRHSSSTRSGGDLTGL 1038
 QY 1070 EPESEBAPSPPL-----APSEGAGSDVPDGDLMGAAGKLGSLPTHDPSPLOYSDDPTV 1124
 DB 1039 -----SRTPLSSLSATSNNSATNCID-----RNGQCHPVREDFVQYSDPTG 1083
 QY 1125 PLPSER-DGYVAPLTSPQPEYVNOQDVRPQPPSPREBPLPAARPAAGTLERAKTSLFG 1182
 DB 1084 NFLLESIDDFL-----PAPRYVNO-LMPKPS-----TAAV 1114
 QY 1183 KNGVVDVF-----AFGAVENPEVLTPOGAAPOPHPPPAPSPAFDNY 1227
 DB 1115 QNGIYNNISLTAISKLPMSORVQNSHSTAVONPEYL-----NTNGSLAKTYFESSP 1166
 QY 1228 YMDQ-----DPE-----KQAPSTFGTPTAENPEVYLGLVP 1260
 DB 1167 YMIQSGNHQINDNDYQODFLPNETKPNGLKVAENPEVYLVAAP 1214

RESULT 7
 A47253
 epidermal growth factor receptor, HER4 - human
 C/Species: Homo sapiens (man)
 C/Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
 C:Accession: A47253
 R:Plowman, G.D.; Culionescu, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
 A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide

A:Reference number: A47253; MUID:93189574; PMID:8363326
 A:Accession: A47253
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1308 <Pro>
 A:Cross-references: GB:L07868; NID:9337359; PIDN:AAB59446.1; PID:9337360
 A:Note: sequence extracted from NCBI backbone (NCBI:126842)
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; protein kinase homology
 F:716-981/Domain: protein kinase homology <kin>
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 43.1%; Score 2945.5; DB 2; Length 1308;
 Best Local Similarity 44.9%; Pred. No. 5,3e-115;
 Matches 609; Conservative 181; Mismatches 378; Indels 189; Gaps 30;

```

9  WGLLLALIPGAA---STOYCTGTMKLRPLSPFPHLMDRLHYGCCOYVQGNLELT 64
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8  WWWVSLVAAAGTVQPSDSQSVCAQTEKLSLSLDEOQYRALRKYENCCEVYMGNLITS 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65 LPTVASLFLDIOEVQGYLIANHOYRQVFLQRLIRYGTQLFEDNYALAVLDNGDPLN 124
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 IEHRDLSFLSVEYVGYVALNQRVYPLENLRIRIKLYEDRYALAIPLNKKDG 127
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 NTPVYTGASPGGLBELQLRSLTEILKGGVLIQRNPQLCYQDTILMKDI FHKNNQALTLI 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 NF-----GLOBELKMLTEILNGVVDQMKFLCYADTIMODIVRNPMPSNLTLLV 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
185 DTRSRACHPCSPMKSGRCWGESSEDCSLTRIVCAGGC-ARGKPLPTDCCHEGCAAG 243
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
179 STNGSSGGRCHKSTG-RCMGPTENHCQTLIRVCAEQCGRCYEPVSSCCRECAAG 237
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
244 CTGPHGSDCLALCFHNSGICELHCPALVTYNTDTFESMPNBSRYTFGASCVTACEPY 303
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 CSGPRDTCFACMNFDSGACVTCQPTFFVYNPTTFQEHNFNAKYTGACVCKCPHNF 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
304 LSTDVSGCTLVCPHNOVTAEDGQRCCKSKCARCYCLMGQYIKANKSFGITBLE 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 V-VDSSSCVRACPSSKMEV-BENGIMKCKPCTDICKPKACDGIIGSLMSACTVDSNIDK 355
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 FAGCKKIFGSLAFLEESFDDPASNTAFLQEQVFTLEITIGYLIASMPSLDLS 423
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
366 FINCTKINGNLIPLVTGIGHDPYNAIEADEKLVNFTVAIEIGFLNIGSPNMTDFS 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 VFQNLQVIRGRILHNGAVSLTQIGISWLGRLSRLREGSLALIHNTHLCPVHTYPMW 483
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 VFSNLYTIGRVLVSGSLILLIKQOGITSLQFOSLKEISAGNIYITDMSNLCYVHTIIMT 475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 QLFRRPHQALLHTARPEDECVGEGLACHOLCARHCGPPTOCVNSOFLROCEVEE 543
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 TLFSTINQRIYIRDKRAENCTAEGMVCNHLCSGCGPDPDCLSRFRSRRGICIES 535
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
366 FINCTKINGNLIPLVTGIGHDPYNAIEADEKLVNFTVAIEIGFLNIGSPNMTDFS 415
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
424 VFQNLQVIRGRILHNGAVSLTQIGISWLGRLSRLREGSLALIHNTHLCPVHTYPMW 483
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
416 VFSNLYTIGRVLVSGSLILLIKQOGITSLQFOSLKEISAGNIYITDMSNLCYVHTIIMT 475
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 QLFRRPHQALLHTARPEDECVGEGLACHOLCARHCGPPTOCVNSOFLROCEVEE 543
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
476 TLFSTINQRIYIRDKRAENCTAEGMVCNHLCSGCGPDPDCLSRFRSRRGICIES 535
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
544 CRVLQGLPREVYNARHCLPCHPECP-ONGSVTCFGEPADOCVACAHYKDPFCVACP- 601
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
536 CNLYGDEFREFENGSIQVECDPCCKMEDGLITLCHGPDPDCTCKSHKCDGPNCEKCPD 595
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
602 ----FNNFTVSFMLRVPKVSASHLEPDEGAQCPRTICTSHCVLDLDDKGC----- 648
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
586 GUGANSTIFKX-----ADPRE--CHPHNCTGCGNGSPITSHDCTYPMWTHS 642
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
649 --PAEORASPLTISAVV-GILLVVLGVVFGILIKRROCKIRKYTRALLQETELVEP 705
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
643 TLPHAR-TPL--IAAGVIGGLIIVGLTFVAVYRRKSIK-KKALRRFL-ELEVEP 697
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
706 LTPSGAMPQAOAMRLIKETELKRYVLSGAFGTYYKGIWPDGNVYKIPAIKYLRENT 765
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
698 LTPSGTAPQOQLRIKETELKRYVLSGAFGTYYKGIWPDGNVYKIPAIKYLRENT 757
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
766 SPKANKELIDAVYVAGVSPYVSRLLGLTSTVQVLTQMPYCCLDLHVRNKGRLS 825
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
758 GKNAVVEFDEALIMASMDHPLVRLVGLVCSPTIQLTQMPHCCLEVEYHKNIDNGS 817
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
826 QLLIMWCQIAGMSYLEDVRLVHRDLAARNVTVKSPNHVKITDGLARLLDIDETETHA 885
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 818 QLLIMWCQIAGMSYLEDVRLVHRDLAARNVTVKSPNHVKITDGLARLLDIDETETHA 877
 QY 886 DGGKPIKMALESLIRRFTHQSDVWSGVTWMLFMFGARPYGCIPIREIPLDLEKE 945
 Db 878 DGGKPIKMALESLIRRFTHQSDVWSGVTWMLFMFGARPYGCIPIREIPLDLEKE 937
 QY 946 RLPPPICTIDYVMIVKCMWIDSECRPFRELVEFSMARDPQRFVYIQMED-LGPAS 1004
 Db 938 RLPPPICTIDYVMIVKCMWIDSECRPFRELVEFSMARDPQRFVYIQMED-LGPAS 997
 QY 1005 PLDSTFRSLLEDDMGDLDVAEEYLVPOQGFCDPAPAGCAWYHRRSSSTSGGSD 1064
 Db 998 PDSKFPQMLDEEDLEDDMDAEEYLV-QAENIPP-----ITSRGARDISNRS---- 1046
 QY 1065 LTLGLEPSEEAPRS-----PLAP-SEGAQDVDFGD 1095
 Db 1047 -ETGSPRAYPMNSGNQVYRDGFAAGQSVVPYRFTSTIPEAPVAGATAEFFDS 1105
 QY 1096 LGWGAAGLQSLPTHDPSFLQYSEDPVPLPS-----ETDGYVAPLTCSPQPEYVQ 1148
 Db 1106 CNGTLRKQVAPRHVQDSSTORYSADPTVAFABERSPGBLDEGYVTPMRDKQOYLNP 1165
 QY 1149 PDVRPQPSBPREGPLPAAPACATLERATLSPGKNGVYKDVAFAGCAVENEYLLPQGG 1208
 Db 1166 VE-----ENPFVSR-----KXGDLQ-----ALDNPYHNSNG 1194
 QY 1209 AAPQPPPPA-----FSPAPDNLYWQDPP 1234
 Db 1195 -----PKRADEYVNEPLYMTFANTLQKAEYLNKNIIMPEKAKKADNDPYWHSILP 1248
 QY 1235 ERGA--PSTFKGTP-----AENPEYL 1255
 Db 1249 PRSTLQHPDYLOEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 8
 506142
 protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
 N:Alternate names: epidermal growth factor receptor homolog; kinase-related transfc
 C:Species: Xiphophorus maculatus (southern platyfish)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
 C:Accession: 506142; S13809
 R:Waltbrod, J.; Adam, D.; Malitchev, B.; Maeueller, W.; Raulf, F.; Telling, A.; R
 Nature 341, 415-421, 1989
 A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing
 A:Reference number: 506142; MUID:90015140; PMID:2797166
 A:Accession: 506142
 A:Molecule type: DNA
 A:Residues: 1-1166 <MIT>
 A:Cross-references: EMBL:X16891; NID:965290; PIDN:CAA34770.1; PID:965291
 R:Adam, D.; Maeueller, W.; Scharf, W.
 Oncogene 6, 73-80, 1991
 A:Title: Transcriptional activation of the melanoma inducing xmrk oncogene in Xiph
 A:Reference number: S13807; MUID:91125882; PMID:1446557
 A:Accession: S13809
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
 A:Cross-references: EMBL:X56319; NID:965284; PIDN:CAA39763.1; PID:965285
 C:Genetics: k
 A:Gene: mrk
 A:Map position: Y
 A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <WAT>
 F:707-972/Domain: protein kinase homology <kin>
 F:725-723/Region: protein kinase ATP-binding motif

Query Match 38.4%; Score 2627; DB 1; Length 1166;
 Best Local Similarity 44.5%; Pred. No. 7,5e-102;

Matches 568; Conservative 165; Mismatches 392; Indels 152; Gaps 31;

```

Qy 4 AALCRWGLLALPPGAASL-----QVCTGDMKRLPLASPEHLDMLRHLYQGQVQGN 59
Db 8 AALQO--LLVLSTSRCCSTDPDRKVGCGTGNQMTM--LDNHLLKKKKKXSGCNVLEN 62
Qy 60 LELTYLPTNASLFLQDIQEVGYVLLAHNOVROVPLQRLAIVRGTOLEFEDNALAVLDN 119
Db 63 LEITYTOENQDLSEFQSIQEVGYVLLAHNEVSTIPIVNLRLIRGQMLYEGNEFTLLWSN 122
Qy 120 GDPINNTPTVGASPGGLREQLRLSEFLEIKGVLIOENPOLQVODITLMDIDHKANQL 179
Db 123 YQK-NPSP--DYVQVGLKQQLSNITHEILSGYKXVSNPLLCVETITNMMDYDKSNP 179
Qy 180 ALTLIDNRSRACHPCSPMCKGRSCWGESSEDCSLRTVYAGGC-ARCKGPLPTDCHE 238
Db 180 TMLNIPHAFERQCKCKDHGCVNCSGMAFPGHCCKFTKLCAEQCNCRCKGPKPIDCCNE 239
Qy 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYMTDFESNPREGRTFGASCUTA 298
Db 240 HGAAGCTGPRATDCLACDFDDGTCOTCPKXITIDVSHQVVDNENIKITFGAACVKE 299
Qy 299 CPYNYLSTDVGSCTLVCPLEHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIG 358
Db 300 CPNRYVTE-GACVRSAGMLLEV-D-ENGKRSCKPCDGVCPKXVCGDIGISL-SNTIAVN 356
Qy 359 ITEL-EPAGCKKITGSLAFLEPSTFDGPASNTAFLQELQVETLEITGYLYTSMPD 417
Db 357 STNRSFSCNCKINGDILINRNSFEGDPHYKIGTMDDEHMLNLTVEITGYLYIMWPE 416
Qy 418 SLPLSVFQNTQVIRGRILHNGAYS-LTLQGLGTSWELSGRLSEGLALIHNTILCF 476
Db 417 NMTLSVFNQLEIRGRITFRSGSFVYVYRHQMLGRLSKVSGNVLKNTLQRY 476
Qy 477 VHTVPMDOLFENPHQALHTANRPDEBCEVGLACHQICARHCWGFQPTQVNCQFLR 536
Db 477 ANTIMRRLFRSEQDSIEYDART-----ENQTCNNECSDEGCM-DGPTMCVSCILHYDR 528
Qy 537 GQECVEGRVYLOGPREYVNAHRLPCPECQPNQSVTCGPEADQVCAHAKXDPFC 596
Db 529 GGRCVASCNLQGEPRQVNDGRVQCHQELVDTSDITCGPPANCSAHQDDQPC 588
Qy 597 VARP---FNNFTVSFMLRVKVASHLPEDEBACOPCPINCTHSCVDDDKCPAECR 653
Db 589 IPRCPHGLDGDTLWKXA-----DKMGCCQPCQHCNCTQGSGPGLSCRD-1 637
Qy 654 ASPSTVSAYVGLLVVGVGGLIKRQCKIRKXTMRLLQELVEYLPPTSGAMP 713
Db 638 VSHSLAVGLVSGLLITVALLVLLRRRIK-RKRTIRCLQEKELVEPLTPSGQAP 696
Qy 714 NQAQMRILKETELRKVYLGSGAFGYKGIWPDGEVNIKPAIKYLRNTSPKAKEI 773
Db 697 NQAFRLIKETEFKDRYLGSGAFGYKGLMNDGENIRIPAIKYLRETSKNOEV 756
Qy 774 LDEAYVWAGVSPYVSRLLGICLTSTVQLVQLMPYGGCLDHYENRGLSGQDLNWC 833
Db 757 LDEAYVWASVDHPVCRLLGICLTSAVOLVQLMPYGGCLDHYOHQERICGQMLNWC 816
Qy 834 QIAKMSYLEDVRLVHRDLARVNLVKSPPNHVKTIDPGLARLIDITEVYHADGKVPK 893
Db 817 QIAKMYLEERHLVHRDLARVNLKPNHVKITTDGLSKLLADEKEVQADGKVPK 876
Qy 894 WMALESILRRRFTHQSDVSYGVTVWELMTFGAKPYDIPAREIPDLLEKGERLLPDPIC 953
Db 877 WMALESILQWYTHQSDVSYGVTVWELMTFGSKPYDIPAKEIASYLENGERLLPDPIC 936
Qy 954 TIDVYMIWKCWMTDSECRPRFRELVESESMARDPQFVYIQWEDGSPASPLDSTYRS 1013
Db 937 TIEVYMIWKWMTDPSRRPRFRELVEFSQMAIDPSRYLTQC--NLPSLSRRLFSR 993
Qy 1014 LLEDDMDGLVDAEYVLPQGFCCPDPAFGAGVHHRHRSSTRSGGDLTLGLEPSE 1073
Db 994 LLSDD--DVVDADEYLLPYKRI-----NRQGS----- 1019

```

RESULT 9

A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; 159164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86: 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epider-
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todarik
Proc. Natl. Acad. Sci. U.S.A. 87: 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor
A:Reference number: 159164; MUID:90311312; PMID:2164210
A:Accession: 159164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:9183990; PIDN:AAA5979.1; PID:9306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase I
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <Kin>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 34.8%; Score 2378.5; DB 2; Length 1342;
Best Local Similarity 39.8%; Pred. No. 1.7e-91;
Matches 526; Conservative 193; Mismatches 460; Indels 141; Gaps 34;

```

Qy 10 GLLDALLPPGA--STQVCTGDMKRLPLASPEHLDMLRHLYQGQVQGNLELYLPT 67
Db 11 GLLFSLARSGVNGSAGVCPGLNGLSVGDANQVQLYKLYERREVMGNLELYLTH 70
Qy 68 NNSLSTLODIQEVGYVLLAHNOVROVPLQRLAIVRGTOLEFEDNALAVLDNDDPINTT 127
Db 71 NADLSFLQWIRETVGVLLVAMNEFSLPLPNIRYVRGTVGYDGKFAIFVM-----LNVNT 125
Qy 128 PVTGASPGGLREQLRLSEFLEIKGVLIOENPOLQVODITLMDIDHKANQLALTLIDN 187
Db 126 ----NSHSLRQLRLQLLEILLSSGYIRKNDKLCMDITIDMDIYRBD---AEIVKD 178
Qy 188 RSBACHPCSPMCKGRSCWGESSEDCSLRTVYAGGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPCHQEVCKG-RCWGPGSBDQTLTKTICAPQCNHCGFPMNQCCHDECAAGCGG 237
Qy 247 PKHSDCLACHFNHSGICELHCPALVTYMTDFESNPREGRTFGASCUTACPYNYLST 306

```

```

Db      238 PODTDCAFACRHFNDGACVPCRPQPLVYNNKLTFOLEPNHPTKYQYGVACVASCPIHFV-V 296
Qy      307 DVGSCITVCPPLHNOEYTAEDGTORCEKSKPCARCYGIMQYIANKSF--IGITLE- 363
Db      297 DDTSCVACRCPDMEVD-KHGLKMEPCGGLCPKACEGTG-----SSRFQTVDSNDIG 350
Qy      364 FAGCKKIFGSLAFLESPDGDASNTAPLOEQLOVFTLEETITGLYISAMPDLPDL 423
Db      351 FVNCSTKIIGNLDFLTGLNGDPMHKKIPALDEPKLVNFTVREITGLYINQSPHMHNS 410
Qy      424 VFQNLQVIRGRILHNGAYS-LTLQGLISWLGRLRELGSGLALIHNTLCFPHYTPW 482
Db      411 VSNLTLTGGRSLVNRGFLIMKLNNTSLGFRSLKEISAGRIYISANROLCTYHSLNW 470
Qy      483 DQLEFNPOALHTA-NRPEDECVGEGLAHQLCARHCWGGPPTQCVNCSQFLRGCECV 541
Db      471 TVLNGPTEERLDIKHNRPRRDCVAEGKCDPLSSGGCGWGGPQGLSCSNYSRGVCV 530
Qy      542 EECRVYQCLPREYVNAHRLPCHPECOPONGSVTFCPEADOCVACAHYKDPFCVACP 601
Db      531 THCNFLNGRPREFAHAEFCFCHPCQPEGTATNGSGSDTCAQCAHFRDGPCHVSCP 590
Qy      602 FNNFTVFWLRVPKYSASHLEPDEGACQPCPINCTSHSCVDLDDKGCFAEQRA----SPL 657
Db      591 HGVLAGK-----GPYKYPDVQNECRPCHENCTQCKGELDDCGLTLVLIGKTHL 642
Qy      658 TSIVSAVGLILVVLGVVFGILIRROOKTR-KYWRRLQETELVPLTPSGAMPNOA 716
Db      643 TMTLVIAQ--LVVIFMMLGFTFLYMRGRIRKRAMRRLYERGESIPPLDPS-EKAKV 699
Qy      717 QWRILKETELRKVKVLGSGAFGTYYKGIMIPDGENVKIPVMIKYLRENTSPKANELLDE 776
Db      700 LARIFKEITELRKVLGSGVFTYKGVWIPGESIKIPVCIKYIEDSGSQSFOAVDTH 759
Qy      777 AYVMAVGSPVYSRLIGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSODLWMCQIA 836
Db      760 MLAGSLDHAHIVRLGLCPGSSLTQVLYPLGSLDHRVORHAGALPQILLMWGVQIA 819
Qy      837 KMSYLEDVRLVHRLAARNVLYKSPHVKITPDELARLLDDEHYHADGKVPYIKMA 896
Db      820 KMYTLEHGVNHRMLAANVLLKSPVOVADPFOVADLPDDQLLYSBAKPIKMA 879
Qy      897 LESILRRFTHOSDVMSYGVTVWELMTGAKPYDIPAREIPDLLEKGERLPQPICTID 956
Db      880 LESIFGKYTHQSDVMSYGVTVWELMTGAEYALRLAEPVDLLEKGERLAQPICTID 939
Qy      957 VYMIWKKCMIDSECRPRRELVSFESKMAADPQRFVYIQNEDLCRA---SPLDSTFYS 1013
Db      940 YVMVWKKCMIDENIRTEKELANFTKARDPRRYLYIKES-GRGIAFGPEPHGLTNK 998
Qy      1014 LLEDDMDGLVDAEYLVPOQGFPCPDPAFGAGKWHHRSSSTRSGGDLTGLP-S 1072
Db      999 KLEEYLEBELDLDDLEAED-----NLATTIGSALSIPVGLTN 1039
Qy      1073 EEEARSPPLABEGAGSDVFCDDLQMGAKGLQSLPTHD-BSPLQRYSEDPVLP- 1127
Db      1040 RPRGOSLISPSG-Y-MPMNQNLGESCOCESAVSSSRCPRPVSLH-----PMRAGCL 1092
Qy      1128 --SETDGYVA-----PLTSPQPE-----YNNQPRVVRQPPSPRGP- 1162
Db      1093 ASESEGHVTSGEALQEKVSMCRSRSRSPRRPGDSAHYSORHSLTTPVPLSPGLE 1152
Qy      1163 -----LPAARPAATLERAKTLP--GKNGV-----KDVFAFGAVENPEYLPQG 1207
Db      1153 EEDVNGVMPDTHLKGTPSSREGTLLSVGLSVLTEREDED-----DEYEMRRR 1204
Qy      1208 GAAPQHPREPAPAFNDLYYWD-----QDPRGAPSTSTKGPRTANPEYL 1255
Db      1205 RHSP-PHPRPSSLEELGYEYVDVGSDSLASLIGSTQSCPLHPVIMPTAGTTPDEDEYEM 1263

```

RESULT 10
JC4387

```

epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1996
C:Accession: J04387
C:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant pro
Gene 165, 279-284, 1995
A:Reference number: J04387; MIM:609653; PMID:8522190
A:Accession: J04387
A:Molecule type: mRNA
A:Rebides: 1-133 <HEL>
A:Cross-reference: GB:U29339; NID:9915389; PID:9915390
A:Experimental source: liver
A>Note: The authors translated the codon AAC for residue 369 as Thr and GTT for re
C:Comment: This protein is a functional heregulin receptor that transduces signals
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase I
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prot
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:160-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1259,1259,1286,1325/Binding site: phosphate (T
Query Match 33.5%; Score 2293.5; DB 2; Length 1339;
Best Local Similarity 40.1%; Pred. No. 5,7e-88;
Matches 518; Conservative 168; Mismatches 440; Indels 165; Gaps 35;
Qy      3 LAALCRMGILLALLPQCA--STVCTGTDMKRLRPASETHLDMIRLYQCCQVYQGN 59
Db      7 LQVLC---FLSLARSEWNSQAVCFGLTDLSTGDADNOYQTLKYKCEVVMGN 62
Qy      60 LELTYPTNASLSFLOIOGVGYVLIANQVAVPQRIYRGQLFEDNVALVLON 119
Db      63 LEIVTGHANDLSFLQIRVTVYLVANNEFVPLNLRVVRGQVVDGKFAIFVM-- 120
Qy      120 GDLPLNTPTPYTASFGGLRELRSLTEILKGGVLIQRNPOLCYODTILMKDIFHNQNL 179
Db      121 ---LNVT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDITDWRDLYVR- 170
Qy      180 ALLIDIDNRBARHPCSPMKCKSGRSGESSEDDQSLTRYVCAGG-ARCKGRLPTOCHE 238
Db      171 GAEIVVKNNGANCPGHEVCKG-RWGPGDDQILLKTIQAQCKGRCGGRPNOCCHD 229
Qy      239 QCAAGCTGPKHSDCLACLNHNSGICELHCRPALVTYNTDTFESMPNREGYTFGASCTA 298
Db      230 ECAAGGCGPQDTCFACRRRNDGACVPRPRELVYNNKLTFOLEPRPHNTKYQYGVACV 289
Qy      299 CFYNYLSTDVGSCITVCPPLHNOEYTAEDGTORCEKSKPCARCYGL--GMQYIANKSKF 356
Db      290 CFHNFV-VDTQFCVACRCPDMEVD-KHGLKMEPCGGLCPKACEGTGSSRYQTVDSN 347
Qy      357 IGITLEFAFCKKIFGSLAFLESPDGDASNTAPLOEQLOVFTLEETITGLYISAMP 416
Db      348 ID-----GFVNCSTKIIGNLDFLTGLNGDPMHKKIPALDEPKLVNFTVREITGLYINQSP 403
Qy      417 DSLPLDSVFQNLQVIRGRILHNGAYS-LTLQGLISWLGRLRELGSGLALIHNTLC 475
Db      404 PHMHNFVSFENLTLTIGRSLVNRGFLIMKLNNTSLGFRSLKEISAGRIYISANROLCTYH 463
Qy      476 FVHTVMPDQCFRPHQALLHTA-NRPEDECVGEGLAHQLCARHCWGGPPTQCVNCSQ 534
Db      464 YHSLNMTKRLRGPSEERLDIKYDRPLGELGAKVCDPLSSGGCGWGPQGLSCSNYSRGVCV 523
Qy      535 LRQCEVCECRVYQCLPREYVNAHRLPCHPECOPONGSVTFCPEADOCVACAHYKDP 594
Db      524 SRGVCVTHCNFLNGRPREFAHAEFCFCHPELCPMEGSTYVNGSSDCAQCAHFRDGP 583
Qy      595 FCVACRPFNNFTVFWLRVPKYSASHLEPDEBACQPCPINCTSHG--VDLDDKGCFAEQ 652

```

Db 584 HCVNSCPHGLGAK-----GPTKKPDAQNECRPHENCTOQCNPELQDCLGQAEV 635

QY 653 RASPLTSTVSASVAGLLVVLGVVFGILLIKRQOKIR-KYTWRRLLQETLVEPLTPSGA 711

Db 636 LMSKPHLVAIVLVG--LAVIIMILIGSFLYWRGRIRQKRAMRRLERGESIEPLDPS-E 692

QY 712 MPNQAKMILKTELKRLKVLGSGAFVYKGIWIPDGENKIPALIVLRENSPRANK 771

Db 693 KANKVLARIFKTELKRLKVLGSGVFGVHGMIPBESIKIPVCIKVIDKSGRQSFQ 752

QY 772 EILDEAYVAGVSGPVYRLLIGLITSTVQLVTLQMLPGCLLDHYRENGRLSGODLLNW 831

Db 753 AVTDHMLAVGSLDHAHIVRLGLCFGSSGLQVLTQVLPGLSLDHYKQRETLGQOLLNW 812

QY 832 CMQIAKMSLYLEDVAVLRDLAARVVLKSPENHYKITDFGLARLIDIDETVHADGKVP 891

Db 813 GVQIAKGMVYLEESHSMVHRDLARVMKSPSOVAVDPLPPDDKQLLSEAKTP 872

QY 892 IKMMLAESILRRRFTHSGDWSYGVTVWELMTFGAKPYDGI-PAREIPPLEKGRLPQPP 951

Db 873 IKMMLAESIHFGKITHSGDWSYGVTVWELMTFGAEFYAGLRALRLEPDLKGRLPQPP 932

QY 952 ICTIDVYIMVYKCMWIDSECRPRRELVSSESRMAQDFQRFVITQNEDLGPASPLDSTFY 1011

Db 933 ICTIDVYIMVYKCMWIDENIRPTEKELANETRMARDPRLVLIKRA-S-GGTP--PAE 989

QY 1012 RSLLEDDMDGLVDAEEYLVPOGFCFDPAPAGAGMWHHRSSSTRSGGDLTLGEP 1071

Db 990 PSVLTTEL-----QEALPEL-----DLDLEA 1015

QY 1072 SEE-----EAPRPLAPSE-----AGSDVFDGLMGAKG 1103

Db 1016 EEEGLATSLGASALPTGTLTRPGSGQLTSPSSGYPMNQSISGEACLDVAVGREQF 1075

QY 1104 LQSLPTHPPLQRYSEDPVPLPSETDGYV----AP-----TC-----SQPE--- 1144

Db 1076 SRPISLH-PIPRGR-----PASESEGHVTGSEAEIQEKVSYCRSRSPRSPRGDS 1127

QY 1145 -YVNOQDVPQPPSPREP-----LPAARPAQATLERAKTISL-P-GKNVY---- 1187

Db 1128 AYHQRHSLTLTPVPLSPGLSEEDGNGYVMDTHMGASSREGTLLSVGLSTGTEE 1187

QY 1188 -KVFAFGAVENPEVLTPOGGAPOPPHP 1216

Db 1188 EDED-----EEVEYNNRKRGRSP-PRPP 1209

RESULT 11

TVFVY

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

C/Species: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C/Species: avian leukosis virus, ALV

C/Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999

C/Accession: B00643 #00643

R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M

Cell 41, 719-726, 1985

A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and p21

A/Reference number: A00643; MUID:85228222; PMID:2988784

A/Accession: B00643

A/Molecule type: mRNA

A/Residues: 1-698 <NTL>

A/Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750

A/Note: in Genbank entry CHKERBPF, release 109.0, the source is designated as Gallus gall

C/Comment: This protein is synthesized as a gag-env-erbB protein.

C/Genetics:

A/Genes: gag-env-erbB

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P

F.1.6/Product: gag protein (fragment) #status predicted <GAG>

F.1.59/Product: env protein (fragment) #status predicted <ENV>

F.60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>

F.194-459/Domain: protein kinase homology <KIN>

F.202-210/Region: protein kinase ATP-binding motif

F.229/Active site: lys #status predicted

Query Match 25.2%; Score 1725.5; DB 1; Length 698;

Best Local Similarity 51.4%; Pred. No. 1e-64;

Matches 373; Conservative 76; Mismatches 138; Indels 139; Gaps 20;

QY 578 GPEADQVCAAHYKDPFCVACRPF-----NNFVSTWLRVVKVYASHLEDEGACQPP 633

Db 60 GP--DHCMKCAHFDIGHCHCVKACPAVLGENDTL-TWKYA-----DANAVALCLH 106

QY 634 INCHSVDLDDDGCPAEQASPLTSTVSASV-VGILLVVLGVVFGILLIKRQOKIRKYT 692

Db 107 PNCTRGCKGGLGECF---NGSKTPSIAGVAGGLCLVVGIGLGLYLR--HIYKRT 162

QY 693 MRLLQETLVEPLTPSGAMPNQAKMILKTELKRLKVLGSGAFVYKGIWIPDGENV 752

Db 163 LRRLQRELVEPLTPSGEAPNQAHRIKTEFFKVKVYLGSGAFVYKGIWIPSEKV 222

QY 753 KIVVALKVLRENSPRANKIILDEAVVAGVSGPVYRLLIGLITSTVQLVTLQMLPGCL 812

Db 223 KIVVALKEIREATSPRANKIILDEAVVAGVSGPVYRLLIGLITSTVQLVTLQMLPGCL 282

QY 813 LDHVRNGRLSGODLLNMQIAKMSYLEDVAVLRDLAARVVLKSPENHYKITDFGL 872

Db 283 LDVIRKHONIGSQVILNMCVQIAKGMVLEERLVRDLAARVVLKTPQHYKITDFGL 342

QY 873 ARLLIDETVHADGKVPVKMMLAESILRRRFTHSGDWSYGVTVWELMTFGAKPYDGI 932

Db 343 AKLLGADKEHYAEGCKVPVKMMLAESILHRIYTHSGDWSYGVTVWELMTFGSKPYDGI 402

QY 933 PAREIPPLEKGRLPQPPCTIDVYIMVYKCMWIDSECRPRRELVSSESRMAQDFQRF 992

Db 403 PASEISVLEKGRLPQPPCTIDVYIMVYKCMWIDADSRPRRELIAESKARDPFRY 462

QY 993 VWIQ-NEDLGPASPLDSTFYRSLLEDDMDGLVDAEEYLVPOGFCFDPAPAGAGMWH 1051

Db 463 LVIQGDEHMLPPTDSKFRITMEEDMEDIDAEYLVPHGFF----- 508

QY 1052 RHSSSTRSGGDLTLGLEPSEBAPRSL-----APSGAGSDVFDGLMGMAKGLQS 1106

Db 509 -NSPST-----SRTPLLSLSLSTSNMSATNCID-----RNGQG 540

QY 1107 LPHDPSPLQRYSEDPVPLPSET--DGVVAPLTCGPOPEYVNOQPPVPPRPGPLP 1164

Db 541 HPVREDSFYQRSSDPTGNGFLBESIDGFL-----PAPSYNQ--LMPKRS----- 585

QY 1165 AARPAQATLERAKTISPGKNGYKVF-----AFGAVENPEVLTPOGGA 1209

Db 586 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL----- 625

QY 1210 APOPHPPPAFSAFDNLVYWDQ-----DPPE-----RGAPSTKGTPTTANPEY 1254

Db 626 -NTNQSPLAKTVFESSPWIGSNHQLINDPDDYQDPLPNETKENGILLKVFAAENPEY 683

QY 1255 LGLDVP 1260

Db 684 LRYAAP 689

RESULT 12

TVVUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain

C/Species: avian erythroblastosis virus

C/Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999

C/Accession: A00644; A58022

R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.

Cell 35, 71-78, 1983

A/Title: The erbB gene of avian erythroblastosis virus is a member of the src gene

A/Reference number: A00644; MUID:84026539; PMID:6313229

A/Accession: A00644

A/Molecule type: DNA

A/Residues: 1-604 <YAM>

A/Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678

R:Debutre, B.; Henry, C.; Benalissa, M.; Biserte, G.; Clavierie, J.M.; Saulle, S.; Martin, Science 224, 1456-1459, 1984
 A:Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of A:Reference number: A38022; MUID:84223357; PMID:6328658
 A:Accession: A38022
 A:Molecule type: DNA
 A:Residues: 1-28, 'W', '30-139', 'F', '141-145', 'V', '147-152' <DEB>
 A:Cross-references: GB:K02006
 C:Genetics:
 A:Gene: erba
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 F:130-355/Domain: protein kinase homology <KIN>
 F:138-146/Region: protein kinase ATP-binding motif
 F:165/Active site: Lys #status predicted

Query Match 24.5% Score 1677; DB 1; Length 604;
 Best Local Similarity 51.6%; Pred. No. 9, 3e-63;
 Matches 361; Conservative 73; Mismatches 126; Indels 140; Gaps 18;

587 CAHYKDPFCVACRCP---NNFTVSFWLRVVKVASHLEPDEGACQCPDINCTHSCVD 642
 3 CAHFIDGPHCVACAGVLGENDTL---VRKVA-----DANAVCQLCHPNCRTGCKG 51
 643 LDDKCGPAPQASPLTISVAVY-GILLVVUGVVFGLIKRROCKIRKYMRLQETE 701
 52 PCLGCP---NSKTPSIAAGVGGGLVWGLGTGLYLKR-HYKRTLRRLQERE 107
 702 LVEPLTPSGAMPNQAOMILKELTKRVKVLGSAFGTVYKGIWIPDENYKIVAIKVL 761
 108 LVEPLTPSGEAPNQHILKELTEFKRVKVLGSAFGITVYKGLWPEGEKXKIVAIKEL 167
 762 RENTSPKANKELIDAVYMAAGSPYVSRLLGICITSTVQLVQLMPYGCILLVHREKRG 821
 168 REATSPKANKELIDAVYMAASVDNHCVRLLGICITSTVQLVQLMPYGCILLVIREKD 227
 822 RUGSODLLNMCQIAKGSYLEVDVLRHDLAARNVLVKSPNHVKTDFGLARLLDIDET 881
 228 NIGSQYLNCQVIAKGNVYLERLVRDLAARNVLVKTHQXKLTDFGLAKLLGADER 287
 882 EYHADGKVPITKMALESILRRRFTHQSDWSYGVTTWELMTFGAKPYDGIIPADEIDLL 941
 288 EYHAGGKVPITKMALESILRRYTHQSDWSYGVTTWELMTFGSKPYDGIIPASEISSVL 347
 942 EKEGRLPQPICTIVYIMVYKWMIDSECPRELFSEFSBMAKRDQRFVIO-NEDL 1000
 348 EKEGRLPQPICTIVYIMVYKWMIDSECPRELFSEFSBMAKRDQRFVIO-NEDL 1000
 1001 GPASPLDSTFYRSLIEDDMGDLVDAEYLVVPOGFCPPDAPAGAGMWHHRHSSSTRS 1060
 408 HLPSPDTSKFRYTLMEEDMEDIVDAEYLVPHQGF-----NSPST-- 449
 1061 GGGDLTLGLEPSEEPSPPL-----APSEGAGSVFPGDGLMGAKLQSLPHHDSPL 1115
 450 -----SRPLLSLSLSTSNNSATNCIT-----KNGQGHVPREDSFV 485
 1116 QRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNVQPDVFPQPSPREGRLPAPAPAGATL 1173
 486 QRYSSDPTGNFLSESIDDGL-----PAPEYVNG--LMPKKPSIAM----- 524
 1174 ERAKTLSPGKNGVVDVFAF-----GGAVENPEVLTQGGAAAPDPH 1214
 555 -----VQNGIYNFISLTAISKLPMSDRYQNSHSTAVNDPEYL-----NTN 564
 1215 PPPAFSPADNLVYWDODPPERKAGPSTFKGTPTAENPEY 1254
 565 QSPLAKTVPESSPYVIOGSHQ-----INLDNDY 594

RESULT 13
 GQFPE
 epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbb

C:Species: Drosophila melanogaster
 C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
 C:Accession: A00640; A38021
 R:Umbh, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z. Cell 40, 599-607, 1985
 A:Title: The Drosophila EGF receptor gene homology: conservation of both hormone bit
 A:Reference number: A00640; MUID:85124611; PMID:2982499
 A:Accession: A00640
 A:Molecule type: DNA
 A:Residues: 1-1330 <LIV>
 A:Cross-references: EMBL:K03054
 R:Madsworth, S.C.; Vincent III, W.S.; Blodgett-Wentworth, D. Nature 314, 178-180, 1985
 A:Title: A Drosophila genomic sequence with homology to human epidermal growth fact
 A:Reference number: A38021; MUID:85137938; PMID:2983332
 A:Accession: A38021
 A:Molecule type: DNA
 A:Residues: 'A', '832-866', 'V', '868-943', 'QTSIVK' <MAD>
 A:Cross-references: EMBL:X02293; MUID:97922; PIDN:CAA26157.1; PID:929565
 C:Comment: This sequence is tentative because the introns have not been identified
 C:Genetics:
 A:Gene: FlyBase:Egfr
 A:Cross-references: FlyBase:FBgn0003731
 A:Map position: 2.57F
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; p
 F:1-732/Domain: extracellular #status predicted <EXT>
 F:732-764/Domain: transmembrane #status predicted <TM>
 F:765-1330/Domain: intracellular #status predicted <INT>
 F:808-1072/Domain: protein kinase homology <KIN>
 F:816-824/Region: protein kinase ATP-binding motif
 F:122,300,324,363,518,668,695,700/Binding site: carbohydrate (asn) (covalent) #stat
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic
 F:843/Active site: Lys #status predicted
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status p

Query Match 23.8% Score 1625.5; DB 1; Length 1330;
 Best Local Similarity 29.6%; Pred. No. 2, 7e-60;
 Matches 412; Conservative 177; Mismatches 414; Indels 387; Gaps 40;

80 VQGVYLIHNVQVQVQLRIVRGTLF-----EDNYALAVLDNGDPNNTPTVYGASD 134
 38 ITNYIVIGDLIPCTLSYRLQIRGRITLSLSEKVALFV-----TV 81
 135 GGLRELQRLTLLKGVYIQNPOLQVDTLWMDIFHKNNQALLTLIDTRSPACH 194
 82 SKMYTLEIPDLRVLVNGQGFHNNYVLCNRTIQWSEIVSNGTDAVYNDFTAPERECP 141
 195 CSPMCKGSRGWSSEDCSLTRTVACGCA--RCGKPLPTDCHEQCAAGCTGPRHSDC 252
 142 CHSCSHG--CWGSGPRNCKQFSKLTSPQACAGRCVGPAPRECHLFCAGGCTGPRQKDC 200
 253 LACIHNHSGICLCPALVTYNTDFESNPNEEGYTTGASCVTACPNYVLTVDGSGCT 312
 201 IACKNFDEAVSEEBPMAKYNPTTYVETNEBGRVAAAGATVKECP--GHILRDGACV 259
 313 LVCPILNNOETADGTCORCEKSPCARCYGQWYIYKANSKFIGITEL-----EPAG 366
 260 RSCFPQKMDKGE-----CYPNGPPEPKTC-----PEVTLHAAGNLSFEN 300
 367 CKKIPLGLAFLESFDC--DPAANTA-----PLQEQQLQVFTELEITGYIYISAMPDS 418
 301 CTVIDNIRILDTDFGQFDVYANVMGPRYIPLDERREVEFTVEIKGYLNIESTHPQ 360
 419 LPLSVFONLOVIRGILHNGAV--STLGLGIGSWGLSLRPLSGLLIHNHTLGV 477
 361 FRNLSTFRNLLETTHGQLMESFPAALAIYKSLYSLEMRNLKQISGSSVVIQHRDLCTV 420
 478 HTVPMQLEFRNPQALLHTANREDEC----- 504
 421 SNIRWPAIQEPEQKWNENLRAD--CGHFLITLISVQNIITHIFAICREKNNHLLGSV 480
 505 ----- 504

Db 481 QGRRLGSMHGSVPYQLDELQFQHLRLMLYIQVINSTQDKSNEHQLTDCYSPSVPT 540
 QY 505 -----VG 506
 Db 541 SLTIERARYAIGSAGLAMELEQITASASMRHSKTLPAERQVPRWVFLGVCASARAGIA 600
 QY 507 EGLA-----CQGLARGHGWGPGTQCVNCSQFLRGQCEVECRVLQGLPREVY--N 556
 Db 601 EPLAGRAVCRKCHPFLCELCTNYGYEHQVCSKCHYGRKQCECTEC-----PADHYTDEE 654
 QY 557 ARRCLEPCHEPCQPNQSVTCFGEADQCVACAHYK-----DPPF-----CVARCFPN 603
 Db 655 QRECFQRHEPC---NG---CTGPGADDCKSCNPFLLPANENGPVYNSTMFCSTKCPLE 708
 QY 604 NPTVSWLWPRKYSASHLEPDEBGACQPCPINCTHSCVDLDKQCPAERASPLTISYSA 663
 Db 709 -----MRHVNQYVAIGP---YCAASPPRSSKITAVLD-----VNMIFI 744
 QY 664 VVGIILVVLGVVFGI-LIKRRQOKIRKYT--MRLLQETELVEPLTPSGAMPNOAMRI 720
 Db 745 ITGAVLVPTICLCVTVICRQKQAKKETVMTALSGREDESEPLRBNIGANLCIKRI 804
 QY 721 LKTELKRVKVLGSGAFGTVYKGIWIPDGENYKIPVALKVLRENTSPKANKIIDEAYVM 780
 Db 805 VVDABELKRGVLMGAFGRVYKGVWVPEGENYKIPVALKELKSGAESSEFLEAYIM 864
 QY 781 AGVSPVYVRLIGICTSTVOLVTOIMPYGCLLDHYRENGRLSGODLLNMCQIAKMS 840
 Db 865 ASDEHYNLKLAVCSQMLITQMLPGLCLDYVRNRRDKIGSKALNMSYQIAKMS 924
 QY 841 YLEDAVRLVHRLAARVLYK---SPNHYKITDFGLARLIDIDETEYHADGKVPRIKMAL 897
 Db 925 YLEERKLVHRDLAARVLYRLAGEDH---DFGLAKLLSDSNHYKAAAGCMPKMLAL 980
 QY 898 ESILRRRPHQSDWSYGTVMELMTFGAKPYDGIIPAREIPDLKKEGRLPOPPICITDV 957
 Db 981 ECIARRVFTSKSDWAFGVTITWELLTFQGRPHENIPAKDIPDLIEVGLKEDEPETSIDI 1040
 QY 958 YVIMVCKMWIDSECRPRELIVSEFSRMAPDQPRFVIVQNEIDLG--PASPLDSTYRSL 1015
 Db 1041 YCTLLSCMHLDAAMPTFKQLTVAEFARDGCRVLALIGDKFTLLPA-----YTQD 1093
 QY 1016 EDD---DMGDLVDAEYLVPOQGFPCPDPAFGAGVHHRRSSSTRSGGDLTLGLEPS 1072
 Db 1094 EKDLIRKLAPTTDGEALAKPDDYLQPKALGPS-----HRTDCT----- 1133
 QY 1073 EEEAP-----RSLAPSGAGSDVFDG---DLGMAKAGLQSLPTHPSPLQRYSEDP 1123
 Db 1134 -DEMKLNRYCKDPENKXISSTDDERDSAREVGVGNLR----- 1171
 QY 1124 VPLPSETDGYAPLTCSPOPEYVNOPIVPPSPREGPLPAPAPAGATLERAKTLSPGK 1183
 Db 1172 LDLPVDEDDYLMF--TCQPGPNNNNNN-----NPNQNMMAVGVAAAYM----- 1214
 QY 1164 NGVVDVAFAGGAVENPEYL---TPQGAAPQPI-----PPAFSP 1221
 Db 1215 -----DLIGVPSVNPPEYLLNAQTLGVGESPITQITIGPYMGPGTMEYKVPWGSEP 1269
 QY 1222 -AFDNLYYWD 1230
 Db 1270 TSSDHEYYND 1279
 RESULT 14
 S35745
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
 C/Species: avian erythroblastosis virus
 C/Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C/Accession: S35745
 R/Vennstrom, B.
 Submitted to the EMBL Data Library, March 1993
 A/Reference number: S35743

A/Accession: S35745
 A/Molecule type: DNA
 A/Residues: 1-544 <VEN>
 A/Cross-references: EMBL.X12707
 C/Genetics:
 A/Gene: erbB
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-spec
 F135-400/Domain: protein kinase homology <KIN>
 F143-151/Region: protein kinase ATP-binding motif
 F170/Active site: Lys #status predicted

Query March 23.5% Score 1606; DB 2; Length 544;
 Best local similarity 53.9%; Pred. No. 7.3e-60;
 Matches 344; Conservative 66; Mismatches 122; Indels 106; Gaps 17;

QY 578 GPEADQCVACAHYDPPFCVACRPF-----NFTVSFWLRYVPKVSASHLEPDEGACQPCP 613
 Db 1 GP--DHCKKCHFIIDGPHCVACAPAGVAGENDTL-VKRYA-----DANVCQLCH 47
 QY 634 INCHSCVDLDKCCPAERASPLTGSVAVY-GILVVVLGVVFGILLKRRQOKIRKYT 692
 Db 48 PNCTRGCKGPGLEQCP---NGSKTPSIAAGVVGILCLVYVVGIGIGLYRRR-HIVKRT 103
 QY 693 MRRLQETELVEPLTPSGAMPNOAMRIKTELKRVKVLGSGAFGTVYKGIWIPDGENV 752
 Db 104 LRLLORELVPEPLTPSGEAPNOAHILKETEKKVVLGFGAFGTVYKGMIPSEKIV 163
 QY 753 KIPVALKVLRENTSPKANKIIDEAYVMAGVSPVYVRLIGICTSTVOLVTOIMPYGC 812
 Db 164 TIPVALKELREATSPKANKIIDEAYVMASVDNPHVCRLLIGICTSTVOLVTOIMPYGC 223
 QY 813 LDHYRENGRLSGODLLNMCQIAKMSYLEDVRLVHRDLAARVLYKSPNHYKITDFGL 872
 Db 224 LDYIREKNDINGSYLLNMCVQIAKGNVYLEERHVRDLAARVLYKTPQHVXITDFGL 283
 QY 873 ARLLIDETEYHADGKVPRIKMALESILRRRPHQSDWSYGTVMELMTFGAKPYDGI 932
 Db 284 AKQAGADEKHYHAGVPIKMALESILRRRPHQSDWSYGTVMELMTFGAKPYDGI 343
 QY 933 PARTIPPLEKEGRLPOPPICITDVIMVCKMWIDSECRPRELIVSEFSRMAPDQPRF 992
 Db 344 PASEISSVLEKGERLPQPPICITDVIMVCKMWIDSECRPRELIVSEFSRMAPDQPRF 403
 QY 993 VVIO-NEDLPASPLDSTFPRSLLEDQMDGLVDAEYLVPOQGFPCPDPAFGAGVHH 1051
 Db 404 LVIOGDERMALPSTDSKFRITLMEEDMEDIVADDEYLVPHQGF----- 449
 QY 1052 RHRSSSTRSGGDLTLGLEPSEEPAPSPPL-----APSEGAGSDVFDGDLGMAKAGLQ 1106
 Db 450 --NSPST-----SRTPLSLSATSNNSATNCIDRNG----- 480
 QY 1107 LPTHDPSPLORYSEDPVPLPSETDGYAPLTCSPOPEYVNOPIVPPSPREGPLPAP 1166
 Db 481 --H-----PVREDDGL-----PAPEYVNO--LMPKPKSTAMVQNIY 513
 QY 1167 RPAGAT-LEBAKTLSPKNGVAVDFAFGAVENPEYL 1203
 Db 514 NYSUTRAISKLPIDSRQN-----SHSTAVNPEYL 544

RESULT 15
 S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
 C/Species: avian erythroblastosis virus
 C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C/Accession: S00727
 R/Scotting, F.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A/Title: Common site of mutation in the erbB gene of avian erythroblastosis virus
 A/Reference number: S00727; MUID:88217326; PMID:2897102
 A/Accession: S00727
 A/Molecule type: DNA

A:Residues: 1-545 <SC0>
A:Cross-references: EMBL:X06943

C:Genetics:

A/Gene: erbB

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; phosphotransferase

F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif

Query Match

23.4%; Score 1599; DB 2; Length 545;

Best Local Similarity 53.9%; Pred. No. 1,4e-59;
Matches 344; Conservative 65; Mismatches 123; Indels 106; Gaps 17;

```
QY 578 GPEADQCVACAHYKDPFCVAPCP-----NFTVSFWLVRPKVASHLPEDEGACQPCP 633
DB 1 GP--DHQMKCAHIDPCHCKACAPAGLVGENDTL-VKXA-----DANAVQQLCH 47
QY 634 INCTHSCVLDLDKGCAPAEORASPLTISAVV-GILLVVVLGVFGILIKRQOKIRKKT 692
DB 48 PNCTRGCKGPGLEGCP---NGSKTPSIAGVVGILCLVVGIGLGLYLRR-HIVRKRT 103
QY 693 MRRLQETELVEPLTPSGAMPNOAKRIKETELRKVKVLSGAFGTYYKGIWPDGENV 752
DB 104 LRRLQERELVEPLTPSGAPNOAHRLKETEFKVKVLSGAFGTYYKGLMPEGEKV 163
QY 753 KIPVAIKVLENTSPANKELIDEAYVMAGVSPVSRLLGICLTSTVQVLTQMPYGL 812
DB 164 TIPVAIKELREATSPANKELIDEAYVMASVDNPHVCRLLGICLTSTVQVLTQMPYGL 223
QY 813 LDHVRNRGRLSGODLLNMCQIAKMSYLEPRLVHRDLAARNVLYKSPNHKLTDEGL 872
DB 224 LDYIRKNDNISQYLLNMCVQIAKMNYLEERHLVHRDLAARNVLYKTPQDKITDFGL 283
QY 873 ARLLDIDETEVHADGKVPDKMALESILRRPFTHSDVMSYGVTVWELMTFGAKPYDGI 932
DB 284 AKQLGADKEKYEHAEGKVPDKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI 343
QY 933 PAREIPDLLEKGRLLPQPICTIDVYMIWVKCMIDSECRPRRELVSFSRMARDPQRF 992
DB 344 PASEISSVLEKGRLLPQPICTIDVYMIWVKCMIDSECRPRRELVSFSRMARDPQRF 403
QY 993 VVIQ-NEDLGPASPLDSTFYRSLLEDMDGLVDAEYLVPOGFCPPDAPAGGMVH 1051
DB 404 LVIQGDERMHLPSPTSKFYRLMEEDMEDIVDAEYLVPHQGF----- 449
QY 1052 RHRSSSTRSGGGLTLGLEPSEEARPSFL---APSEGAGSDVDFGDLGMGAAGLQS 1106
DB 450 --NSPST-----SRTPLSLSLSATSNNSATNCIDRNG----- 480
QY 1107 LPTHDPSPLORYSEDPVLPSETDGVVAPLTCSPQEVYNOQDVPPSPPREGLPAA 1166
DB 481 ---H-----PVREDGFL-----PAPEYVNC--LMPKRPSTAMVQNCIY 513
QY 1167 RPAGAT-LEPAKTLSPKRGVVKDVFAGGAVENPEYL 1203
DB 514 NYISLTAISKLPMDSRYN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:10:21
Job time : 32.0157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(Without alignments)
5088.033 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 6841

Sequence: 1 MEUAALCRMGCLLALPPGA.....TEKQPTAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6621	96.8	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5833	85.3	1257	1 ERB2_RAT	P06494 rattus norv
3	5817.5	85.0	1254	1 ERB2_MESAU	P00553 mesocricetu
4	3108	45.4	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3080	45.0	1210	1 EGFR_MOUSE	P01279 mus musculu
6	2945.5	43.1	1308	1 ERB4_HUMAN	P15303 homo sapien
7	2922	42.8	1308	1 ERB4_RAT	P02986 rattus norv
8	2683.5	38.6	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2387.5	34.9	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2316.5	33.9	1339	1 ERB3_RAT	P062799 rattus norv
11	1960	28.7	1426	1 EGFR_DROME	P04412 drosophila
12	1708.5	25.0	634	1 ERBB_AYER	P00534 avian leuko
13	1677	24.5	604	1 ERBB_AYER	P00535 avian eryth
14	1589	23.2	540	1 ERBB_AYER	P11273 avian eryth
15	1554	22.7	703	1 EGFR_CHICK	P13387 gallus gall
16	1305	19.1	1323	1 LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.7	245	1 ERB2_MOUSE	P70442 mus musculu
18	730	10.7	1363	1 ILPR_BRLA	O02466 branchiosto
19	710	10.4	1300	1 IRR_MOUSE	O09114 mus musculu
20	695	10.2	1297	1 IRR_HUMAN	P14616 homo sapien
21	694	10.1	1382	1 IRR_HUMAN	P06223 homo sapien
22	694	10.1	1383	1 IRR_HUMAN	P15127 rattus norv
23	693	10.1	1477	1 HTK_HYDAT	Q25197 hydra atten
24	691.5	10.1	1372	1 INSR_MOUSE	Q25198 mus musculu
25	691	10.1	1607	1 MIPR_LYMT	Q25410 lymphocytic
26	688.5	10.1	1300	1 IRR_CAVPO	P14617 cavia porce
27	647	9.5	1367	1 IGR_HUMAN	P08069 homo sapien
28	629	9.2	1373	1 IGR_MOUSE	P08071 mus musculu
29	625.5	9.1	1370	1 IGR_RAT	P24062 rattus norv
30	620	9.1	1390	1 INSR_AEDAE	Q93105 aedes aegy
31	614	9.0	2146	1 INSR_DROME	P09208 drosophila
32	597	8.7	987	1 EPB4_HUMAN	P54760 homo sapien
33	594.5	8.7	977	1 EPB4_MOUSE	O03145 mus musculu

34	591.5	8.6	984	1 EPB1_RAT	P09759 rattus norv
35	588.5	8.6	976	1 EPB2_HUMAN	P29317 homo sapien
36	586.5	8.6	984	1 EPB1_CHICK	O07494 gallus gall
37	585.5	8.6	984	1 EPB1_HUMAN	P54762 homo sapien
38	583	8.5	902	1 EPB3_XENLA	O91736 xenopus lae
39	581.5	8.5	985	1 EPB4_XENLA	O91571 xenopus lae
40	577	8.4	987	1 EPB4_MOUSE	P54761 mus musculu
41	574.5	8.4	1053	1 FAK1_CHICK	O00944 gallus gall
42	574	8.4	1114	1 RET_HUMAN	P07749 homo sapien
43	569	8.3	1068	1 FAK1_XENLA	O91738 xenopus lae
44	566	8.3	757	1 HT16_HYDAT	P53356 hydra atten
45	563	8.2	1052	1 FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234 (1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Gauseens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGarrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139 (1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=8616729; PubMed=2995967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429 (1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP20 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- P1M: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.ebi.ac.uk/announcements/](http://www.ebi.ac.uk/announcements)
 CC or send an email to license@ebi.ac.uk).
 CC
 DR EMBL; M11767; AAA5808.1; JOINED.
 DR EMBL; M11761; AAA5808.1; JOINED.
 DR EMBL; M11762; AAA5808.1; JOINED.
 DR EMBL; M11763; AAA5808.1; JOINED.
 DR EMBL; M11764; AAA5808.1; JOINED.
 DR EMBL; M11765; AAA5808.1; JOINED.
 DR EMBL; M11766; AAA5808.1; JOINED.
 DR EMBL; M11730; AAA5808.1; JOINED.
 DR EMBL; M12036; AAA5808.1; JOINED.
 DR EMBL; X03363; CAA27060.1; JOINED.
 DR PIR; A25491; A25491.
 DR PIR; A24571; A24571.
 DR HSSP; F11362; 1FGK.
 DR GeneW; HGNC:3430; ERBB2.
 DR MIM; 164870; ERBB2.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00261; Fu; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 KW
 FT SIGNAL 1 21
 FT CHAIN 22 1255
 FT DOMAIN 22 652
 FT TRANSMEM 653 675
 FT DOMAIN 676 1255
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT DISULFID 195 204
 FT DISULFID 199 212
 FT DISULFID 220 227
 FT DISULFID 224 235
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 531 540
 FT DISULFID 544 560

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 654 654 I -> V.
 FT VARIANT 654 654 /FTID=VAR_004077.
 FT VARIANT 655 655 I -> V.
 FT CONFLICT 1170 1170 /FTID=VAR_004078.
 FT SEQUENCE 1255 AA; 137909 MM; 399DFDA04DF962 CRC64;
 SQ
 Query Match 96.8%; Score 6621; DB 1; Length 1255;
 Best Local Similarity 96.8%; Pred. No. 0; Mismatches 19; Indels 14; Gaps 2;
 Matches 1225; Conservativity 7;
 QY 1 MELIALCRWGLTALPPGAASVCTGTDMKRLPASPETHLDMRLHYOGQVQGNL 60
 DB 1 MELIALCRWGLTALPPGAASVCTGTDMKRLPASPETHLDMRLHYOGQVQGNL 60
 QY 61 ELTYLPPTNASLSTLQIOEVQGVLIANQVQVPLQRLRYRGTLQFEDNYALVLDNG 120
 DB 61 ELTYLPPTNASLSTLQIOEVQGVLIANQVQVPLQRLRYRGTLQFEDNYALVLDNG 120
 QY 121 DPLNNTPTVPGASPGGLRELQRLSTELIKGGLIQRNPOLQCYDTILMKDIFHKNNOLA 180
 DB 121 DPLNNTPTVPGASPGGLRELQRLSTELIKGGLIQRNPOLQCYDTILMKDIFHKNNOLA 180
 QY 121 DPLNNTPTVPGASPGGLRELQRLSTELIKGGLIQRNPOLQCYDTILMKDIFHKNNOLA 180
 DB 121 DPLNNTPTVPGASPGGLRELQRLSTELIKGGLIQRNPOLQCYDTILMKDIFHKNNOLA 180
 QY 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACARCKGLPTDCHEQC 240
 DB 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACARCKGLPTDCHEQC 240
 QY 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACARCKGLPTDCHEQC 240
 DB 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACARCKGLPTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTPESMNPGRYTFGASCYACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTPESMNPGRYTFGASCYACP 300
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTPESMNPGRYTFGASCYACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTPESMNPGRYTFGASCYACP 300
 QY 301 YNYLSTVGSCTLVCPLEHNEVTAEDGTQRCCKSPKARVCYGLQWYIKANSKFIGIT 360
 DB 301 YNYLSTVGSCTLVCPLEHNEVTAEDGTQRCCKSPKARVCYGLQWYIKANSKFIGIT 360
 QY 301 YNYLSTVGSCTLVCPLEHNEVTAEDGTQRCCKSPKARVCYGLQWYIKANSKFIGIT 360
 DB 301 YNYLSTVGSCTLVCPLEHNEVTAEDGTQRCCKSPKARVCYGLQWYIKANSKFIGIT 360
 QY 361 ELEPAGCKTIFGSLAPLPSFGDDPASNATAPLOPQLOVETLEETIGYIYIAMPDLSL 420
 DB 361 ELEPAGCKTIFGSLAPLPSFGDDPASNATAPLOPQLOVETLEETIGYIYIAMPDLSL 420
 QY 361 IOEPAGCKTIFGSLAPLPSFGDDPASNATAPLOPQLOVETLEETIGYIYIAMPDLSL 420
 DB 361 IOEPAGCKTIFGSLAPLPSFGDDPASNATAPLOPQLOVETLEETIGYIYIAMPDLSL 420
 QY 421 DLSVFOQLVYIRGILHNGAYSLTLOGLISWLGSLSLRGLALIHNTLTCVHTV 480
 DB 421 DLSVFOQLVYIRGILHNGAYSLTLOGLISWLGSLSLRGLALIHNTLTCVHTV 480
 QY 421 DLSVFOQLVYIRGILHNGAYSLTLOGLISWLGSLSLRGLALIHNTLTCVHTV 480
 DB 421 DLSVFOQLVYIRGILHNGAYSLTLOGLISWLGSLSLRGLALIHNTLTCVHTV 480
 QY 481 PMQDLFNPQALLTANRDEDCVGEGLACHQLCARGHGCGFPQCVNCSQFLRQEC 540
 DB 481 PMQDLFNPQALLTANRDEDCVGEGLACHQLCARGHGCGFPQCVNCSQFLRQEC 540
 QY 481 PMQDLFNPQALLTANRDEDCVGEGLACHQLCARGHGCGFPQCVNCSQFLRQEC 540
 DB 481 PMQDLFNPQALLTANRDEDCVGEGLACHQLCARGHGCGFPQCVNCSQFLRQEC 540
 QY 541 VEEPRVYQGLPREVNARHCLPCHPEQOPNGSVTGGPADCCVACAHKDPFCVARC 600
 DB 541 VEEPRVYQGLPREVNARHCLPCHPEQOPNGSVTGGPADCCVACAHKDPFCVARC 600
 QY 541 VEEPRVYQGLPREVNARHCLPCHPEQOPNGSVTGGPADCCVACAHKDPFCVARC 600
 DB 541 VEEPRVYQGLPREVNARHCLPCHPEQOPNGSVTGGPADCCVACAHKDPFCVARC 600
 QY 601 PFNNFTVFWLRYPKVSASHLE---PDEGACQPCPINCTHSCVDLDDGCPAEGRASP 656
 DB 601 PSG-----VKPDLSTYMPIMKPFDEGACQPCPINCTHSCVDLDDGCPAEGRASP 650
 QY 657 LTSIVSANGVGLLVVYGVVFGILLIKRQCKIRKYMRLLOSTELVEPLTPSGAMPNCA 716
 DB 657 LTSIVSANGVGLLVVYGVVFGILLIKRQCKIRKYMRLLOSTELVEPLTPSGAMPNCA 710
 QY 657 LTSIVSANGVGLLVVYGVVFGILLIKRQCKIRKYMRLLOSTELVEPLTPSGAMPNCA 710
 DB 657 LTSIVSANGVGLLVVYGVVFGILLIKRQCKIRKYMRLLOSTELVEPLTPSGAMPNCA 710

FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.
 FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.3%; Score 5833; DB 1; Length 1257;

Best Local Similarity 85.4%; Pred. No. 5,4e-305;

Matches 1082; Conservative 53; Mismatches 116; Indels 16; Gaps 4;

QY 1 MELALCRWGLLALPPGAASVCTGTDMKRLPASPEHLDMLHLYOGCQVQGNL 60
 DB 1 MELAAWCGFLALPPGIAGTGTCTDMKRLPASPEHLDMLHLYOGCQVQGNL 60
 QY 61 ELTYLPTNASLPIODIOEVGYVLIANOVROYPLRLRIVRGSTOLFEDNYALAVLDNG 120
 DB 61 ELTYVPAVASLPIODIOEVGYVLIANOVKRYVPLQRLRIVRGSTOLFEDNYALAVLDNR 120
 QY 121 DPLNNTTPEVT-GASPGJLRELQSLTEILKGVLIQRPOLCYQDTILMKDIFHKNNOL 179
 DB 121 DPGDVAVASTPGRTPEGLRELQSLTEILKGVLIQRPOLCYQDVMWLMKDVFRKNNOL 180
 QY 121 DPGDVAVASTPGRTPEGLRELQSLTEILKGVLIQRPOLCYQDVMWLMKDVFRKNNOL 180
 QY 180 ALTLIDTRSRACHPCSPMKGSKGSSSEDQSLRTVAGCAGCCKGLPPTDCHQ 239
 DB 181 APVIDITNRSRACHPCSPMKGSKGSSSEDQSLRTVAGCAGCCKGLPPTDCHQ 240
 QY 240 CAAGCTGPKHSDCLACHFNHSGICECHCPALVYNTDFESMNPREGRYFGASCYTAC 299
 DB 241 CAAGCTGPKHSDCLACHFNHSGICECHCPALVYNTDFESMNPREGRYFGASCYTTC 300
 QY 300 PYNVISTVSGCTLVCLAHNOEVTADGTQCEKSKPCARAVCYGLGMQYIKANSKRTGI 359
 DB 301 PYNVISTVSGCTLVCPNNQOEVTADGTQCEKSKPCARAVCYGLGMHILRGARATSD 360
 QY 360 TELEFAGCKTIFGSLAFIPESFDGDPASNTAPLOPELOQVETLEBEITGYIYISAMPDSL 419
 DB 361 NVQEPDCKKIFGSLAFIPESFDGDPASNTAPLOPELOQVETLEBEITGYIYISAMPDSL 420
 QY 420 PDLVSFQNLQVIRGILHNGAVSLTGLGISWLGNSRLSGGLAIHNNTHLCFVHT 479
 DB 421 RDLVSFQNLRIIRGILHNGAVSLTGLGISHLGNSRLSGGLAIHNNTHLCFVHT 480
 QY 480 VPMOGLFENPHOALLHTANRPEDE-CVGEGLACHQLCARGCMGPGPTOCNCSQPLRG 538
 DB 481 VPMOGLFENPHOALLHTANRPEDE-CVGEGLACHQLCARGCMGPGPTOCNCSQPLRG 540
 QY 539 ECVEECRVLQGLPREYVNAHRCPLCHPECPQNSVTCFGPADQCVACAYKXDPFCVA 598
 DB 541 ECVEECRVWKGIPREYVNAHRCPLCHPECPQNSVTCFGPADQCVACAYKXSSCVA 600
 QY 599 RCPENFTVSWFLARPKXSASHLE---PDEEGACQCPINICTSHCVLDLXKGPAPQRA 654
 DB 601 RCPSC-----VKPDLSTMPWKYPDEEGICQCPINICTSHCVLDLXKGPAPQRA 650
 QY 655 SPLTIVSAVVGILLVVLGVVFGILLIRROOKIRKXTRMLLOTELVEBLVETSGAMPN 714
 DB 651 SPLTIVSAVVGILLVVLGVVFGILLIRROOKIRKXTRMLLOTELVEBLVETSGAMPN 710
 QY 715 QAQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIVALKVLRNTSPKANKEL 774

DB 711 QAQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIVALKVLRNTSPKANKEL 770
 QY 775 DEAYVAGVGSFVYSRLGICLTSVQVLTQMLPYGCLLDHYRENGRGLSDLLNMCQ 834
 DB 771 DEAYVAGVGSFVYSRLGICLTSVQVLTQMLPYGCLLDHYRENGRGLSDLLNMCQ 830
 QY 835 IAKGMSYLEDVPLVHRDLAARNVLYKSPHNVITPFGRLRLDIDETEXHAOGKVPICM 894
 DB 831 IAKGMSYLEDVPLVHRDLAARNVLYKSPHNVITPFGRLRLDIDETEXHAOGKVPICM 890
 QY 895 MALESILRRRPTHQSDVMSYGVTELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954
 DB 891 MALESILRRRPTHQSDVMSYGVTELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 950
 QY 955 IDVYIMVYKCMWIDSCRRPRELVSFESRMARDPQRFVIONEDLGPSSPMDSTFYRSL 1014
 DB 951 IDVYIMVYKCMWIDSCRRPRELVSFESRMARDPQRFVIONEDLGPSSPMDSTFYRSL 1010
 QY 1015 LEDDDMGDLVDAEYLVPOQGFPCDPAPAGAMVHRRSSSTRSGGDLTLGLEPSEE 1074
 DB 1011 LEDDDMGDLVDAEYLVPOQGFPCDPAPAGAMVHRRSSSTRSGGDLTLGLEPSEE 1070
 QY 1075 BAPRSLAPSEAGSDVFPDGLCMGAKLGLPTHDSEPLQYSEDPIVPLPSETDGY 1134
 DB 1071 BAPRSLAPSEAGSDVFPDGLCMGAKLGLPTHDSEPLQYSEDPIVPLPSETDGY 1130
 QY 1135 APITGSPQPEYVYVQPPRPPSPREGPLPAARPAATLERAKTISPGKNGVYKDVFAFG 1194
 DB 1131 APITGSPQPEYVYVQPPRPPSPREGPLPAARPAATLERAKTISPGKNGVYKDVFAFG 1190
 QY 1195 GAVENBEYLTPOGGAAPQHPPPAPFASPAFNDLYWDDPDERGAPPESTKGTPTANPEY 1254
 DB 1191 GAVENBEYLTPOGGAAPQHPPPAPFASPAFNDLYWDDPDERGAPPESTKGTPTANPEY 1250
 QY 1255 LGIDVPV 1261
 DB 1251 LGIDVPV 1257

RESULT 3
 ERB2_MESAU STANDARD; PRT; 1254 AA.
 ID ERB2_MESAU
 AC Q60553;
 DT 15-DEC-1998 (Rel. 37, last sequence update)
 DT 15-DEC-1998 (Rel. 37, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (neu proto-oncogene) (C-erbB-2).
 GN ERB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_Taxid=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T., Tahijima T., Iehikawa T.,
 RA Yamazaki Y., Iehikawa T.,
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene."
 RL Gene 140:251-255(1994).
 CC -|- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -|- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -|- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

RESIDUES.

- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: D16295; BAA03601.1; -

HSSP: P11352; 1FCX.

InterPro: IPR000494; EGF_L_domain.

InterPro: IPR000719; Euk_pkinase.

InterPro: IPR002174; Furin-like.

InterPro: IPR001245; Tyr_pkinase.

InterPro: IPR004019; YLP_mocif.

pfam: PF00069; pkinase; 1.

pfam: PF00757; Furin-like; 1.

pfam: PF01030; Recep_L_domain; 2.

ProDom: PD000001; Euk_pkinase; 1.

SMART: SM00261; FU; 3.

SMART: SM00219; Tyrc; 1.

PROSITE: PS00107; PROTEIN KINASE ATP; 1.

PROSITE: PS00109; PROTEIN KINASE_TYR; 1.

PROSITE: PS50011; PROTEIN KINASE_DOM; 1.

Transmembrane; Glycoprotein; Multigene family; Receptor; Signal; Transference; Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene; Disease mutation.

CHAIN 1 21

DOMAIN 22 1254

TRANSMEM 653 675

DOMAIN 676 1254

DOMAIN 158 368

DOMAIN 472 644

NP_BIND 720 987

BINDING 753 753

ACT_SITE 845 845

DISULFID 195 204

DISULFID 199 212

DISULFID 236 244

DISULFID 240 252

DISULFID 255 264

DISULFID 268 295

DISULFID 299 311

DISULFID 315 331

DISULFID 334 338

DISULFID 511 520

DISULFID 515 528

DISULFID 531 540

DISULFID 544 560

DISULFID 563 576

DISULFID 567 584

DISULFID 587 596

DISULFID 600 623

DISULFID 626 634

DISULFID 630 642

MOD_RES 1139 1139

MOD_RES 1247 1247

CARBOHYD 68 68

CARBOHYD 125 125

CARBOHYD 187 187

CARBOHYD 259 259

CARBOHYD 530 530

CARBOHYD 571 571

CARBOHYD 629 629

CARBOHYD 658 658

VARIANT 659 659

SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.0%; Score 5817.5; DB 1; Length 1254;
 Best Local Similarity 85.0%; Pred. No. 3-6e-304;
 Matches 1075; Conservative 62; Mismatches 113; Indels 15; Gaps 3;

QY	1	MELALCRWGLLALLPPGAASVCTGTDMKRLRPLASPEHLDMLNHLVGGCVVQGNL	60
DB	1	MELAAWGWGLLALLPPGASGTCTGTDMKRLRPLASPEHLDIVHLVGGCVVQGNL	60
QY	61	ELTLPNNAASLFDIQLQEVQYVLIHANOVRQVPLRLIVRSTQLFENYALAVLDNG	120
DB	61	ELTLPNNAATSLFDIQLQEVQYVLIHANSQVRVPLRLIVRSTQLFEDRYALAVLDNR	120
QY	121	DLPLNNTPVYTGASPGQLRELQLSLTEILKGVLIQRNPOLCYODTILMKDIFKNNOLA	180
DB	121	DLPLNNTVATQRTREGRLRELQLSLTEILKGVLIIRNPOLCYODTILMKDIFKNNOLA	180
QY	181	LTILDTNRSRACHCSFCKSGSRGWSSSDQSLTPTVACGACRKPLPTDCCHQC	240
DB	181	PVDIDTNRSRACPCPCACCKDNHGMGASPEDCOTLTGTIAPRAVPAARARLPTDCCHQC	240
QY	241	AACTGPKHSDCLAFHNSGICELCPALVTNTDTPESMPREGRTFGASCYACP	300
DB	241	AACTGPKHSDCLAFHNSGICELCPALVTNTDTPESMPREGRTFGASCYATCP	300
QY	301	YNYLSTDVSGCTLVCPILHNOEVTAEQDQCEKSKPCARVCYGLGMQYIKANSKFIGIT	360
DB	301	YNYLSTVSGCTLVCPILHNOEVTAEQDQCEKSKPCARVCYGLGMHKGARALITSAN	360
QY	361	ELFENAGCKTFGSLAPLPSFDPDGPANTAPLOEQVETLEETLFGYLYISWEDSLP	420
DB	361	IQEPAGCKTFGSLAPLPSFDPGSSGIALPPEQVETLEETLFGYLYISWEDSLH	420
QY	421	DLSTFQVQLVYRGRILNNGAVSLTQLGLSMGLRSRLSGSLAIHNNTHLCFHTV	480
DB	421	DLSTFQVQLVYRGRVLDHGVSLALQGLRMLGLRSLRSLGSLVLIHNNTHLCFHTV	480
QY	481	PMQDLFNPQIALIHTNRPEDBCVSGELACHQICARHGCGPQTQVNCQSLRQEC	540
DB	481	PMQDLFNPQIALHNSNPSEBECGLKDFACYPICAHGHGCGPQTQVNCQSHLRQEC	540
QY	541	VEEGRVYQGLPREVYVNRHGLCPHPCPOPGNSVTCGPRAQDCVCAHAKDPGCVARC	600
DB	541	VKEGRVYQGLPREVYVNRHGLCPHPCQPNSTETCTGSAQDCTAPPHKDSFCVARC	600
QY	601	PFNNFTVSPWLAVPKVSASHLE----PDEBGAQCPCEINCTHSCVDLDDKGPAPQASP	656
DB	601	PSG-----VKPDLSTMPIMKYPDEBGMQPCINCTHSCVDLDDKGPAPQASP	650
QY	657	LTSIVSARVVGILLVAVGVFGILLIKRRQKIRKYMRLLOETELVEPLTPSGAMPNOA	716
DB	651	ATSLIATVVGILLVAVGVFGILLIKRRQKIRKYMRLLOETELVEPLTPSGAMPNOA	710
QY	717	QMRILKETELRKVYVLSGSAFGTVYKGIWIPDEBNYKIPAIKVLRENTSPKANKELDE	776
DB	711	QMRILKETELRKVYVLSGSAFGTVYKGIWIPDEBNYKIPAIKVLRENTSPKANKELDE	770
QY	777	AYWAGVGSPPYRRLGLICTSTVQLVTQMPYGLLDHVENRGRIGSODLLNMCQIA	836
DB	771	AYWAGVGSPPYRRLGLICTSTVQLVTQMPYGLLDHVENRGRIGSODLLNMCQIA	830
QY	837	KGMSYLEDVAVLRDIAARVAVLKSPNHVYITDPGLARLLIDETVHADGGVPIKMA	896
DB	831	KGMSYLEDVAVLRDIAARVAVLKSPNHVYITDPGLARLLIDETVHADGGVPIKMA	890
QY	897	LESILRRPHTGSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID	956
DB	891	LESILRRPHTGSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID	950
QY	957	VYVIMYKCMWIDSECPREPLVSEFSRMAADQRRVVVIONEDLGASPLDSTFFYSLE	1016
DB	951	VYVIMYKCMWIDSECPREPLVSEFSRMAADQRRVVVIONEDLGASPLDSTFFYSLE	1010

QY 1017 DDDMGDLVDAEYLYPQCGFPDPAAGWVHHRRSSSRSGCGDLTGLPSEBEA 1076
 DB 1011 DDDMGDLVDAEYLYPQCGFPDPAAGWVHHRRSSSRSGCGDLTGLPSEBEA 1070
 QY 1077 PPSPLAPSEAGSDVFDGDLGMAAKGLOSLPTHDPSPLOXYSEDPVLPSETDGYVAP 1136
 DB 1071 PPSPLAPSEAGSDVFDGDLGMAAKGLOSLPTHDPSPLOXYSEDPVLPSETDGYVAP 1130
 QY 1137 LTCSPPEYVNPQDVPAPPPPPREGPLPAPAPAGATLEAKTSLGKGVVADVAFCGA 1196
 DB 1131 LACSPPEYVNPQDVPAPPPPPREGPLPAPAPAGATLEAKTSLGKGVVADVAFCGA 1190
 QY 1197 VENPEYLTPOGGAAPPP 1256
 DB 1191 VENPEYLTPOGGAAPPP 1249
 QY 1257 LDVFPV 1261
 DB 1250 LDVFPV 1254

RESULT 4

EGFR HUMAN

PRT; 1210 AA.

ID EGFR_HUMAN STANDARD; PRT; 1210 AA.

AC P00533; P06268; Q14225; Q9JMD7; Q9JMD8; Q9JMG5; Q92795; Q00732;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor

GN EGFR OR ERBB1)

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI_TaxID=9606;

RN [1] SEQUENCE FROM N.A. (ISOFORM 1).

RX MEDLINE=84219729; PubMed=6328312;

RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,

RA Lee J., Yarden Y., Liberman T.A., Schlessinger J., Downward J.,

RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;

RT "Human epidermal growth factor receptor cDNA sequence and aberrant

RT expression of the amplified gene in A431 epidermoid carcinoma cells";

RL Nature 309:418-425(1984).

RN [2] SEQUENCE FROM N.A. (ISOFORM 2).

RP TISSUE=Placenta;

RX MEDLINE=95382957; PubMed=7654368;

RA Ilekis J.V., Stark B.C., Scoccia B.;

RT "Possible role of variant RNA transcripts in the regulation of

RT epidermal growth factor receptor expression in human placenta";

RL Mol. Reprod. Dev. 41:149-156(1995).

RN [3] SEQUENCE FROM N.A. (ISOFORM 2).

RP TISSUE=Placenta;

RX MEDLINE=97078686; PubMed=8918811;

RA Reiter J.L., Mainle N.J.;

RT "A 1.8 kb alternative transcript from the human epidermal growth

RT factor receptor gene encodes a truncated form of the receptor";

RL Nucleic Acids Res. 24:4050-4056(1996).

RN [4] SEQUENCE FROM N.A. (ISOFORM 2).

RP TISSUE=Placenta;

RX MEDLINE=97256547; PubMed=9103388;

RA Ilekis J.V., Garlet J., Niederberger C., Scoccia B.;

RT "Expression of a truncated epidermal growth factor receptor-like

RT protein (TEGFR) in ovarian cancer";

RL Gynecol. Oncol. 65:36-41(1997).

RN [5] SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).

RP TISSUE=Placenta;

RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schehl Sinclair C., Pearall R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Mainle N.J.;

RT "Comparative genomic sequence analysis and isolation of human and

RT mouse alternative EGFR transcripts encoding truncated receptor

RT isoforms";

RL Genomics 71:1-20(2001).

RN [6] SEQUENCE OF 575-687 FROM N.A.

RP Reiter J.L., Thredgill D.W., Danielson A.J., Schehl C.M.,

RA Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RA Mainle N.J.;

RT "Human and mouse alternative EGFR transcripts encoding only the

RT extracellular domain of the receptor";

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [7] SEQUENCE OF 713-924 FROM N.A.

RX MEDLINE=84196372; PubMed=6326261;

RA Lin C.R., Chen W.S., Krulger W., Stolarsky L.S., Weber W.,

RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;

RT "Expression cloning of human EGF receptor complementary DNA: gene

RT amplification and three related messenger RNA products in A431

RT cells";

RL Science 224:843-848(1984).

RN [8] SEQUENCE OF 150-962 FROM N.A.

RX MEDLINE=84245835; PubMed=6330563;

RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,

RA Roe B.A., Merlino G.T., Pastan I.,

RT "Human epidermal growth factor receptor cDNA is homologous to a

RT variety of RNAs overproduced in A431 carcinoma cells";

RL Nature 309:806-810(1984).

RN [9] SEQUENCE OF 1028-1210 FROM N.A.

RP MEDLINE=85046483; PubMed=6093780;

RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,

RA O'Malley B.W.;

RT "Isolation of an evolutionarily conserved epidermal growth factor

RT receptor cDNA from human A431 carcinoma cells";

RL Biochem. Biophys. Res. Commun. 124:125-132(1984).

RN [10] SEQUENCE OF 1-29 FROM N.A.

RX MEDLINE=88217333; PubMed=3329716;

RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,

RA Waterfield M.D.;

RT "The human EGF receptor gene: structure of the 110 kb locus and

RT identification of sequences regulating its transcription";

RL Oncogene Res. 1:375-396(1987).

RN [11] SEQUENCE OF 1-29 FROM N.A.

RP MEDLINE=91107677; PubMed=1988448;

RA Haley J.D., Waterfield M.D.;

RT "Contributory effects of de novo transcription and premature

RT transcript termination in the regulation of human epidermal growth

RT factor receptor proto-oncogene RNA synthesis";

RL J. Biol. Chem. 266:1746-1753(1991).

RN [12] SEQUENCE OF 1-29 FROM N.A.

RX MEDLINE=85270438; PubMed=2991899;

RA Ishii S., Xu Y.H., Strutton R.H., Roe B.A., Merlino G.T., Pastan I.;

RT "Characterization and sequence of the promoter region of the human

RT epidermal growth factor receptor gene";

RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).

RN [13] SEQUENCE OF 540.

RP Submitted (SEP-1997) to the SWISS-PROT data bank.

RX MEDLINE=84191554; PubMed=6325948;

RA Mroczkowski B., Mosig G., Cohen S.;

RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and

RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-126; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Otake M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=97297456; PubMed=1039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGR, 3/p110 and 4; are produced by
 CC alternative splicing
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X00588; CA25240.1; -
 CC DR EMBL: U95089; AAB51063.1; -
 CC EMBL: U48722; AAC50802.1; -

DR EMBL: U48723; AAC50804.1; -
 DR EMBL: U48724; AAC50796.1; -
 DR EMBL: U48725; AAC50797.1; -
 DR EMBL: U48726; AAC50798.1; -

Query Match 45.4%; Score 3108; DB 1; Length 1210;

Best Local Similarity 49.6%; Pred. No. 46-159;

Matches 633; Conservative 169; Mismatches 351; Indels 124; Gaps 26;

11 LLLALLPQAA--STQVCTGTDMKRLRPSFTHLDMRLHYOGCGVGNLETLFVN 68
 14 LLAALCPARALEEKVCQGTNSKLTQLCTFEEDHLSLQRMNNECVLGNLEITVQKN 73
 69 ASLSFLQDIQEVQGVLLAHQVROVPLQRLRIVEQTQLFEDNYALAVLDNGDPLNTP 128
 74 YLSFLKTIQEVAGVLLALNTVERIPLENLQIIGNMYENSALAVLSND----- 126
 129 VVGASGGI-REIQLSTIELLGVLLIQNPQLCYQDTLMDIFHKNNQLALLTIDTR 188
 127 ---ANKTGKLELPMRLQELHGAVFSSNNPALCVSESIQWRDIVSDFLSNMSMDPQNR 183
 189 SPACHPCSPMCKGRSCWSESSDCSLTRTVAGGCA-RCKGPLPTDCCHECCAGCTGP 247
 184 LSCQCKDPSGSCNGSCWAGEENQKLTITLCAQCCSGRCGRKSPSDCHNCCAGCTGP 243
 248 KASDCLACHFNHSGICEHCPALVTYNTDFESPMPNBEGRYTPGASCTYACPVYLTSD 307
 244 RSDCLVCCKRFDEATCKDTCPLMLYNPTTYQMDVNEBGKYSFATCVKCKPRNVYTD 303
 308 VASCTLVCPLEHNOETAEADGTORCEKSPCARVCYGQMQYIKANSKFIQTELE-FAG 366
 304 HSCYRACADSYEM-EDGVKCKCKCEPCCKVNGGIGSPK-DLSLNTNTHKFNK 361
 367 CKRTGSLAFLESPSDGPASNTAPLQPEQLQVFTLEIRTYLYISAMPDLPPLSVQ 426
 362 CTSISGDHLILVAERGDSFTHTPLDQELDLTKVKEITGFLILQAMPENRTDLHARE 421
 427 NQVIRGRLHNGASLTLQIGISMLGLRSRLREGSLALIHNTLFCFVTVRWDLF 486
 422 NLEIRGRKQKQGLAVSLAVSLNTSLGRLSKLEISDQVITSGKNKCYANTIMKKLE 481
 487 RNPQALLHTANRPDECEVGEGLACHQLCARHGCWGPPTQCVNCSQFLRGCEVEEYV 546
 482 GSGQKTKIISRGNSCKATQVCHALCSPEGCMGPBROCVSGRNYSRGCEYDKNL 541
 547 LOGLPREYNARHCLPCHEPCOPQNGSVTTCGPEADQCVACAHYKDPFCVACRPF 602
 542 LEGEPREFEVENSECTQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCFAGVWG 601
 603 -NNFTVSPFLRVKVSASHLEPDEGACQPCINCTHSCVLDLDCGCPAEQASPLSTIV 661
 602 ENNTLV-W---KVADAGH-----CHLCPKCTYVCTGPGLEGCFTNGPKIP--SIA 647
 662 SAVVG--ILLVVLGVVFGILIKRQOKIRKTYRRLLOETELVEPLTPSGAMNQAM 718
 648 TQMVGALLLLLVVALGIG--LFMRRRHIVKRTLRLLQRELEVEPLTPSGEAVNQLL 704
 719 RLKXTELKRYVVGSGAFGYTYKGIWIPDENYKIPIAIVLEPNTSPYANKETLIDAY 778
 705 RLKTEFKTIVLSSGAFGVYKGLMIPSEKVIPIAIVLEPNTSPYANKETLIDAY 764
 779 VMAVGSPYVSRLLGICLTSTVQVLTQMPYGLCDHRENRGRGLSGODLNMWQIAKG 838
 765 VVASVDNPHVCRLLGICTSTVQVLTQMPYGLCDHRENRGRGLSGODLNMWQIAKG 824
 839 MSYLEDVAVLRDLAARVLYKSPNHVYKITPFGARLLDIDETRYHADGKVPYIKWALE 898
 825 MNYLEDRLVLRDLAARVLYKTPQVYKITPFGAKLGAEEKVHAEGGVPIKMWALE 884
 899 SILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIREIPMLEGGERLPPQPICTIDVY 958
 885 SILHRIYHQSDVMSYGVTVWELMTFGSKPYDGIPIASISITLKEGRLLPPQPICTIDVY 944


```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 69052046DFD2F5 CRC64;

```

Query Match 45.0%; Score 3080; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 1.3e-157;
 Matches 635; Conservative 161; Mismatches 360; Indels 128; Gaps 27;

```

QY 11 LLLALLPPGAA-STOYCTGTDMKRLPASPETHLDMRLHYOGCQVQVGNLTLYLPTN 68
DB 14 LITLCAAGALBEKKYCOGTSNRLTOLGTFEDHFLSQRYNNCEVLLGNLETTYQNRN 73
QY 69 ASLSFLDIOEVOGVVLIHANGVAVLOPLRLIRGQVPELDNALAVLNDGPDLPNTTP 128
DB 74 YDSFLKTIQVAGVYLIATLVNTVERIPLENQITIRGALVYENTYALAILSN----- 124
QY 129 VTGASPGGLRELQRLSTELIKGVLIQRNPOLCYODTILMKDI----FKHNOALATLI 184
DB 125 -YGTNRGTGLRELPMRNLOELIGAVRFSNNPILCNMTIQRDIYQVVENFSMSMDL--- 180
QY 185 DTNRSRACHPCSPKCGSRGWSSEEDCQSLTRTVCGGCA-RCKGPLPTDCHEGCAAG 243
DB 181 -QSHPSSCPKCDPCSPGSCGCGEENCOKLTKLIQAOCSHRGRSPSDCGHNCQCAAG 239
QY 244 CTGPKHSDCLACTLFFNHSIGICELHCPALVTYNTTFESMPNPEGRYFEGASCYACPPNY 303
DB 240 CTGPRESDCLVCQCFQDEBATCKDTCPELMILNPTTYQMDVNPBEKYSFGATCYAKCEPNY 239
QY 304 LSTDVSGCTLVCPHANOVTAEADGTQCEKSKPCAVCYGLQMOTIKANSKEIGITELE 363
DB 300 VVTDHSGCVACGPDYEV-EBDGIKCKKCDGCRKACNGIGIGERK-DTLISNATNIK 357
QY 364 -FAGCKKIFGSLAFPEPQDPASNTAPLOPELOVFEETLEETGLVYISAMPDLPDL 422
DB 358 HFKCTAISGDHLPLFAFKGDSITRPLPDPRELEILKTVKEITGGLLIQAMPDMDTDL 417
QY 423 SVFQNLQVIRGLIHNGAVYLQGLGISWLGRLSELGSLALIHNTLGFVTPV 482
DB 418 HAEENLEIIRGRTKQHGQFSLAVVGNITSLGRSLKEISDGVYIISGNRLTYANTIM 477
QY 483 DQFPRNHQALLHNPANPEDECVGEGLAGHQLCARHGCMRPGTQCNCQCFRGQECVE 542
DB 478 KKLFGTNPQTKTKMNAEKDCAVNHVNCPLSSSEGCWPEPRDVCSCNVSRGREVE 537
QY 543 ECRVLOGLPREYVVARHRLPCHPECOFONGSVTCFPEADQCAVCAHYKDPFCVRCPE 602

```

```

DB 538 KCVILGEPRREFENSECICQCHPECUPQMANNTICTRGDPNCICQCAHYIDGPHCVITCPA 597
QY 603 -----NNFTVSFWLVRPKVASHLPEDEGACOPCINCHSCVDDDKCCPAEGRASPL 657
DB 598 GINGENNTLV--WKYA-----DANNVCHLANCTYGCAGGGLQGCCEVWPSGPKI 645
QY 658 TSIVSAVAGLILVVVGVVFGI-LIKROOKIRKYMRLLOETLEVEPLTPSGAMPNQ 716
DB 646 PSIAITGIVGGLFIV-VALGIGLFMRRHIVKRTLRLOREIVEPLTPSGEAPNQ 704
QY 717 QMLILETELKRVUGSAGFGVYVGINIPDEANKIPIATIVVAENSPKANKELDE 776
DB 705 HRLILKTEKTKIKVIGSAGFGVYVGLMIPDEKXKIPAIKELAEASPANKKELDE 764
QY 777 AYMAGSFPYVSRLLGICITSTVQVLTQMLPAGCLLDHRENGRLGSDLLMCMQOLA 836
DB 765 AYMVAVDNPHVRRLLGICITSTVQVLTQMLPAGCLLDHREKNDIGSYLLMCMQOLA 824
QY 837 KMSSTYEDVRLVHDLAANNVLYKSPNHYKITDFGLARLLIDETEVHADGKVPARKMA 896
DB 825 KGMNVLDEDRVLVHDLAANNVLYKTPQHYKITDFGLAKLLGAEEKEYHAGGKVPARKMA 884
QY 897 LESILRRPFGSDVWSYGVYVWELTFGAKPYDGI PAEIPDLLEKGERLPQPICTID 956
DB 885 LESILIRIYTHOSDWSYGVYVWELTFSKPYDGI PASDISILEKGERLPQPICTID 944
QY 957 VYMIWVKCWMIDSECPREELVSEFSNAPRQFQEVVIO-NEDLGPASPLDSTFYRSLL 1015
DB 945 VYMIWVKCWMIDSDSPKRELELLEFSKMAPOQRLVIOGDERMHLPSPTOSNFRALM 1004
QY 1016 EDDMDGLDVADEEYLVPOQGFPCPDPAAGMWHHRHSSSTRSGGGLITGLSEBEE 1075
DB 1005 DEEDMDVADAEYLVPOQGF-----NSPST----- 1031
QY 1076 APRSPAPSEAGSDVDFDGLGKAAGKQSLPTHPSPLOKRSBEDPTVLPSEET--DGY 1133
DB 1032 -SRTPLLSLSATSN-----NSTYACINRNGSCVYEDALQKYSSTPTGAVEDNIDDAF 1086
QY 1134 VAEPLTSPQPEYVNOVDVFPQPPSPREGFLPAARPGATLEBAKTLISPKNGVYKDVFAF 1193
DB 1087 L-----PVPEYVNO-SVPKRPAVSQVFNHYNQPLHP-----APGRDLHYQN--EH 1129
QY 1194 GGAVENTEYL-TPQGAAPQPPHPPAFSAFNLVYWDQ-----DP-----PER 1236
DB 1130 SNVAGNPEYLVNTAQ-----PTCLSSGFNSPALMIQKSHJWSLDNPDYQODFFPK 1180
QY 1237 GAPPSFTKGTPTAENPEYVLGLDVP 1260
DB 1181 TKPNGIFKG-PTAENNAEYLRVAPF 1203

```

RESULT 6
 ERB4_HUMAN STANDARD; PRT; 1308 AA.
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (P180erbB4) (tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=93189574; PubMed=8383326;
 RA Plozman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 RA Roy L., Neubauer M.G., Shoyab M.,
 RT Ligand-specific activation of HER4/p180erbB4, a fourth member of the

RT epidermal growth factor receptor family";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elettue K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester.";
 RL J. Biol. Chem. 272:26761-26768(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
 CC 2, NRG-3. HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BECAMELLULIN AND
 CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. (shown here) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS.
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L07868; AAB59446.1; -;
 CC HSSP: P11362; 1FGK.
 CC DR Genew; HGNC:3432; ERBB4.
 CC MIM: 600543; -;
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recsep_L_domain; 2.
 DR Pfam: PF02757; YLP_2.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00261; Fur_4.
 DR SMART: SM00219; TyKc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Alternative splicing.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 652 675 POTENTIAL.
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 186 334 CYS-RICH.
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.

FT NP BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843 BY SIMILARITY.
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 BY SIMILARITY.
 FT MOD_RES 1188 1188 BY SIMILARITY.
 FT MOD_RES 1258 1258 BY SIMILARITY.
 FT MOD_RES 1284 1284 BY SIMILARITY.
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 626 648 NGPTSHCIYPMWGHSITLPPQHA -> IGSSIEDICIGLMD
 (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 58AAB0985D88761 CRC64;
 Query Match 43.1%; Score 2945.5; DB 1; Length 1308;
 Best Local Similarity 44.9%; Pred. No. 2.2e-150;
 Matches 609; Conservative 181; Mismatched 378; Indels 189; Gaps 30;
 QY 9 WGLLLALLPQGA-----STGYCTGIDMKLRIPASPEHRLDMRLHYOCQVVOGNETLY 64
 DB 8 WYWSLVLAAGTVPDSOSVCACTENKLSLSLDEQYRLRKTYENCEVYMNLEITS 67
 QY 65 LPTNASISFLDIOEVQYVLIHNOVRQVPLRLRIVRGTLQFEDNYALAVDNGDFLN 124
 DB 68 IEHNRDLFFLSVREVTGYVVALNORFYLPLEMLRIIRGKLYEDRYALAIFFNYKDG 127
 QY 125 NTTPTVTSFGGLRELORSITLILKGVLIQRNPOLCYOITLIMKOIFHNGQLTLTI 184
 DB 128 NF-----GLQELGLKLTLEILNGCVYVDQNKFLCYADITHQDVRNPPASNTLIV 178
 QY 185 DTNRSFACHPSPMKGRCKRGESSEDQSLTRVYCAAGC-ARCKGPLPTDCHQCAAG 243
 DB 179 STNSSGCGRCHKSCG-RCWGPTEHNCQTLTRVYCAAGCGRGVPGVSDCHRECAAG 237
 QY 244 CTGPRHSCCLAHNHSGICELHCPALVYNTOTFESMPRBSGRYTFGASCVAAPPNY 303
 DB 238 CSGPDTGCFACGMFNDSGACVTCQPTQVFNPTTFQLEHFNKAYTYGACVCKCPNF 297
 QY 304 LSTDVSGTLLVCLPHNCEVTAEDGTORCEKSKPCARVYGLAQVIYKANSKIGITELE 363
 DB 298 V-VSSSGVRAKSSKMEV-BENGIKMKPCPTDICKRACDSIGISLMSAQTVSSNIDK 355
 QY 364 FAGCKITGSLAFLEPESGDPASVNPAPLOEQVETLEITIGYLYISAMPDPLSL 423

```

Db      356  FINCTKINGNLIPLVTGHDGPNVAIEAIDEKLNIVRTVEITGFINIOSWPNMTDFS 415
Qy      424  VFCNLQVIRGRILHNGAVSLTLQGLISWLGRLSRLBELGSLALIHNTHLCEFTVPMW 483
Db      416  VFSNLVITIGRVLVLSGLSLILLKQOGLTSLQFQSLKELSAKNITITDNSNLCHYHTIMT 475
Qy      484  QLFENPHQALHTANRPEDCEVGGGLACHQLCARGHCMGPGPTCVNCSQFLRQCEVEE 543
Db      476  TLFSTINQRIYIRDNRAKENCJTAGMVCNHLCSDDGCGPDDCLCSFRFSRGRICIES 535
Qy      544  CRVLOGLPREVYVNAHCLPCHPECOP-ONGSVTCFGEADCCVACAYKDPPEVCVACP- 601
Db      536  CNLVDGEREFENSGICVCECDPCQCBKMEDGLITLHGQPDNCTCSHFKQPCNCEKCPD 595
Qy      602  ---FNNFTVSFWLRVPRKVSASHLEPDEGACQCPFNCTHSCVDLDDKGC----- 648
Db      596  GLQGANSEIFKY-----ADPRE--CHFCNENCTGCGNGPTSHDCIYYFWTGH 642
Qy      649  --PAQRASPLTISAVV--GILLVVLGVVFGILLIKRQOKIRKYMRLLOTELVEP 705
Db      643  TLPOHAR-TPL--IAAGVIGLFLVLTGTFAYVYRRKSTK-KKRLRRPL-ETELVEP 697
Qy      706  LTPSGAMENQOMEILKTELKRVKVLGSGAFGVYKGIWIPDENYKIPVAILKYLENT 765
Db      698  LTPSGTAPNQQLILKTELKRVKVLGSGAFGVYKGIWIPDEETVKIPVAILKILETT 757
Qy      766  SPXANKELDEAYVMAVGSPYVSRLLGICLTISVOLVLTQMPYGCILLHYRENGRLG 825
Db      758  GPKANVEFMDALMASDHDPHVLVLLGVCLSPICVLTQMPHGLLEVEYHEKMDIGS 817
Qy      826  QDLNMGCMQAKGMSYLEDVVLVHRLDAARNVLKSPNHYITDPFGARLLDIDETEYHA 885
Db      818  QLLNMGCMQAKGMSYLEDVVLVHRLDAARNVLKSPNHYITDPFGARLLDIDETEYHA 877
Qy      886  DGVVPIKMALESILRRFTHSDVMSYGVTVWELMTFGAKPYDIPAREIPDLLEKGE 945
Db      878  DGVVPIKMALESILRRFTHSDVMSYGVTVWELMTFGAKPYDIPAREIPDLLEKGE 937
Qy      946  RLPOPICTIDVYIMVCMWIDSECRPREELVSESRNARDQRFVIVQND-LGPAS 1004
Db      938  RLPOPICTIDVYIMVCMWIDSECRPREELVSESRNARDQRFVIVQND-LGPAS 997
Qy      1005  PLDSTFYRSLLEDMDGLDYDAEEYVLPQOFGPCDPAPGAGVHHRRHSSSTRSGGD 1064
Db      998  PNDKFPQNLDEDEEDMDDAEYVLP-QAFNPP-----LYTSRAIDENRS----- 1046
Qy      1065  LTGLPSESEBAPPS-----PLAP-SEGAGSDVFDGD 1095
Db      1047  -EIGHSPPAVTPMSGQFYVYRDGFAEGGVSVYRAPTSITPEAVAGGATAEITFDDS 1105
Qy      1096  LGMGAAGLQSLPFDSPSLQRYSEDPVPLPS-----ETDGYVAPLTCSPQPEYVQ 1148
Db      1106  CCNGLTKRPVAPHQDESDSTQRYADPTVAPERSPRGELDEEYIMPMCDKROEYILN 1165
Qy      1149  PDVFPQPPSPREGGLPAPAPGATLTERAKTLSPGKNGVAVKDAFGAVENPEYLTPOG 1208
Db      1166  VE-----ENPFVSR-----KNGDLO-----ALDNPYHNASNG 1194
Qy      1209  AAPQHPPPA-----FSPAPNLYVMQDDP 1234
Db      1195  ---PPKADEVENPEYLYINTFANTLGAAYIKNNIISMPKAKAFNPDPTVMHSLP 1248
Qy      1235  ERGA--PSTFFKGT-----AENDEYL 1255
Db      1249  PRSTLQHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

```

RESULT 7
 ERB4_RAT
 ID ERB4_RAT STANDARD; PRT; 1308 AA.
 AC Q62956; Q922N7;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN      ERB4 OR TYRO-2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON      NCBI_TaxId=10116;
RX      MEDLINE=98221155; PubMed=9553078;
RC      TISSUE=Heart;
RA      Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulin promotes survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      [2]
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=91222560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system."
RL      Neuron 6:691-704(1991).
RN      [3]
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulin and their putative receptors, Erb2 and
RT      Erb3, is induced during Wallerian degeneration."
RL      J. Neurosci. 17:1642-1659(1997).
CC      -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF- $\beta$ , AND AMPHIREGULIN (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERB
CC      RECEPTORS (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
DR      EMBL; AF041838; AACD08899.1; .
DR      HSSP; U52531; AAC53051.1; .
DR      HSSP; P11362; 1FGK.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_Pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_Pkinase.
DR      InterPro; IPR004019; YLP_motif.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF00069; pkinase; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      Pfam; PF02757; YLP_2_domain; 1.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PD000001; Euk_pkinase; 1.

```

DR SMART: SMO0261; FU: 4.
 DR SMART: SMO0219; TYRC: 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 652 675 POTENTIAL.
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 186 334 CYS-RICH.
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.
 FT NP_BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843 BY SIMILARITY.
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).
 FT CONFLICT 1062 1062 S -> N (IN REF. 3).
 FT CONFLICT 1080 1082 PVT -> SYR (IN REF. 3).
 SQ SEQUENCE 1308 AA; 146957 MW; D944BB0956A08B41 CRC64;

Query Match 42.8%; Score 2329; DB 1; Length 1308;
 Best Local Similarity 44.6%; Pred No. 1.7e-149;
 Matches 606; Conservative 188; Mismatches 388; Indels 176; Gaps 29;

QY 1 MELA-ALCRGILL--ALLPQASTOVCTGDMKLRIPASBETHLDMRLHYOGCCVQV 57
 DB 1 MCLATGLWVGSLVAARTVQPSASQSCAGTENTLSLSLDEQYRALRKYENCEVVM 60

QY 58 GNLLETLYPTNASLFLDIOEVGVYVLIANQVQVPLQRLRYRGTLQEDNVAALAVL 117
 DB 61 GNLLETLSLEHNDLFLRSIRVTVGVVALNQFYPLPLENRIIRGKLYEDRALAIF 120

QY 118 DNGDPLNNTTPTVLTGASFGGLRELQRLSLTEILKSGVLIQRNPOLCYQDTILMKDIFHGN 177
 DB 121 LNYRKGDF-----GLQELGLKNTLEILNGSVYVDQKKFLCYADTIHMODIVRPFM 171

QY 178 QLALTLIDTNRSPACHPCSPMKSGKSCWGESSEDCSLTRTVACAGC-ARCKGPLTDDC 236
 DB 172 PSMNTLVSTIGSSGGRCHKSCG-RCKPTEHNCOTLRITVAEQDCKRCYPRVSDC 230

QY 237 HEQCAAGCTGPKHSDDLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTFGASCV 236
 DB 231 HRECAAGCGSPKTDTCFACNMFNDSGACTQCQPTFVYNTTFQLHFNPAKTYGAFCV 230

QY 297 TACPYYLSTDVSGSLVLCPLHNEQTAEDGTORCKGSKPCARVCYCGMOYIKANSKF 356
 DB 291 KCPHNFV-VDSSCVRACPSCKMEV-EEGIMCKPCPTDICKACDQIGTGLMSAQV 348

QY 357 IGITTEFAAGCKIFGSLAFLEPSPGDDPASNTAPQPEQVETLEETGLYISAMP 416
 DB 349 DSNINIKFICTKINGNLFLVTGIGHDPYNAIDALDEPLXNFRVTRITGLNICTWP 408

QY 417 DSLPDLSPVFNQVITGRILHNAYSLTLQGLGSLWGLSRLSELGSLALHNNHLCF 476
 DB 409 PNMIDSVFENLVITIGRVLVSGSLILIKQGITSLQFSLKEISAGNIYITDNLCLY 468

QY 477 VHTVPMQDLPRNHOALHTANRDECEVGEGLACHQLCARHCWMPGPTQVCNCSQFLR 536
 DB 469 YHTINMTLTFSTYNTQRTVIRDNRAENCTAEGVCHNLCSNDGCMRGPRQCSGRFSR 528

QY 537 GQECVEECRVLQGLPREYVNAHCLPCHPECQ-ONGSYTCFGEADQVCAHYDPPF 535
 DB 529 GKICIESCNLYDEDEFEFENGSLCEVCDSCCEKMEBGLTLCHGPRGDMCTKCHFXDGP 588

QY 596 CVARCP-----FNNFVTSFWLRVPKVSASHLEDEDEGAQCPINCHSCVLDIDGCG-- 648
 DB 589 CVERCPDVLOGANSIFKY-----ADQRECHFCHPNCTQCGNGTSHDCTY 635

QY 649 -----PAQRASPLTISYAVV-GILLVVLGVFGILIKRQOKIRKYMRLQ 698
 DB 636 YPMTGSHLPQHR-TPL-IAAGVIGFLVTALMTFAVYRRKSIR-KKALRRFL- 690

QY 699 ETELVEELTSSGAMPQAOQRLKETELRKVKTLGSGAGTGYKGIWIDGEMVKIPAI 758
 DB 691 ETELVEELTSSGAMPQAOQRLKETELRKVKTLGSGAGTGYKGIWIDGEMVKIPAI 750

QY 759 KVLRENTSPRANKELIDEAVVMAGSPYSRLGICLSTVOLVQLMAYGCLLHVRE 818
 DB 751 KILNETTPRANVEFMDALMASVPHPLVRLVGLCSTLTOLVQLMAYGCLLHVRE 810

QY 819 NRGRLSGQILLNCMQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKTIDFGLARLDI 878
 DB 811 HKDNIGSQLLLNMCVQIAKGMYLEERLVRDLAARNVLKSPNHVKTIDFGLARLDI 870

QY 879 DETEYHADDGKVIKMMALLESILRRFTHOSDVMSGVYWEIMTGAKPYDGIIPAREIP 938
 DB 871 DEKEYNADGSMETKMMALLECTHYRKFTHOSDVMSGVYWEIMTGAKPYDGIIPAREIP 930

QY 939 DLLEKGRLEPQPICTITDVMYWKCMMDISECRPPRELVSFSSMARDPQRFVIONE 998
 DB 931 DLLEKGRLEPQPICTITDVMYWKCMMDISADRPKFEIALEFSSMARDPQRFVIONE 990

QY 999 D-LGPAASPLDSTFYRLLEDMDGDLVDAEVYLVPOQGFPCDP----- 1041
 DB 991 DRMLKLPENDSKFFQNLDEEDLDDMDAEVYLV-DAEFTIPPIYTSRTIDSNSSETG 1049

QY 1042 -----APGAGGVHRRHRSSTRSGGDLTLGLEPSEEARSPPLASSEGADVDFG 1094
 DB 1050 HSPRPATPMSGQFYVQDGFATQCG---MPPRYATITSTIEPAVA--QGATAMFDD 1104

QY 1095 DLGMAKAGLQSLPTHDPSFLQYSEDPTVPLPS-----ETDGVVAPLTGSPQPEYVN 1147
 DB 1105 SCNGNGLRKRVVHVOEDSSQRYSDPVTFAERNPRALDEGGYVTPMHDKPKQKEYLN 1164

QY 1148 QPVRPQPSPPREGPLPAARPAAGATLEAKTSLPGKGVGVQVFAFGAVENPEYLTPOG 1207
 DB 1165 PVE-----ENPVSRR-----KXGDLQ-----ALDNEFYSAS 1193

QY 1208 GAAPQHPHPA-----ESPAPDNLVYWDQDP 1233

Db 1194 G-----PPKAEDEYNEPLLYLTFTNALGNAEYKNSLLSVEKAKKAFDNDYMHSL 1247
 QY 1234 PERGA--PSTFKGTPT-----AENPEYL 1255
 Db 1248 PPSTSLHPDYLOEYKTYKYNKGRIRPIVAENPEYL 1285

RESULT 8
 XMRK_XIPMA STANDARD; PRT; 1167 AA.
 ID XMRK_XIPMA
 AC P13388;
 DT 01-JUN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorphia; Acanthopterygii; Perciformes; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OC NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Meltbrodt J., Adam D., Malitschek B., Maueker W., Raulf F.,
 RA Telling A., Robertson S.M., Schartl M.,
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RT inducing Tu locus in Xiphophorus.";
 RT Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Schartl M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein-
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X16891; CAA34770.2; -
 CC PIR: S06142; S06142.
 CC HSSP: P13362; 1FCG.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002290; Ser_thr_kinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF000669; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SMO0261; FU; 5.
 DR SMART: SMO0220; S_TKc; 1.
 DR SMART: SMO0219; TYKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642 KINASE.
 FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 666 1167 POTENTIAL.
 FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 716 724 PROTEIN KINASE
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835 ATP (BY SIMILARITY).
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 226 BY SIMILARITY.
 FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT DISULFID 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 141 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.6%; Score 2643.5; DB 1; Length 1167;
 Best Local Similarity 44.6%; Pred. No. 2.8e-134;
 Matches 569; Conservative 165; Mismatches 392; Indels 151; Gaps 30;

QY 4 AALCWGILLALPPGAASST---QVCTGTDMKRLPASPETHLDMRLHYOGCOVQGN 59
 Db 8 AALLQ--LLLVLSISRCSTDPDRKVCQGTSTNQM---LDVHYLKMKKMVGSCNVVLEN 62
 QY 60 LELTYLPFNASISFQDIOEVQGYVLIHNOVAVPQLRLRVRGTOLPEDVYALAVLDN 119
 Db 63 LEITTOENODLSFQISIQEVQGYVLIHNEVSTIPLVNRLIRQNTIEGNTFTLVMSN 122
 QY 120 GDPNNTPVTGASFGSLRELQRLSLEILKGVLIQNPOLCYODTILMKDIFKNNQL 179
 Db 123 YQK-PPSSP--DVYVGLKQQLSLVLTETILSGVYVSHNPPLCNVETINMMDVKTSP 179
 QY 180 ALTLIDTMRSPRCHCSMCKGSRMGWSSSSDDQSLTFTVACAGC-AACKGPLPDDCHE 238
 Db 180 TMLILPHAFERQCCQCDHGCYVNGSCWABPGHCQFTYLLAEQGNRCRGPIDCCNE 239
 QY 239 QCAAGCTGPKHSDCLACLFHNSGICELFALVYNTDTFESMPNPGRTTFCASCYTA 298
 Db 240 HCAGGCTPBRATDCLAGCDFNDGCTCKTCTPPKYLIVSHQVNDVNPXIKTFGACAYKE 299
 QY 299 CPYNYLSDVGSCTLVCLHNOETAEQGTORCEKSKPCARVCYGLGMOYIKANSKIG 358
 Db 300 CPSNVVVE-GACVRSASGMLVND-ENGKSKCRPCDVCVKVCDGIGSL-SNTTAVN 356
 QY 359 ITEL-EFAGCKKIFESLFLPESFGDPAASNTAPLPQPOLQVFTLEBITGYLYISAMPD 417
 Db 357 STNIRSFNCIKINDIILNNSPBGDHYIKIGMDPEHMLNLTIVKEITGYLYVAMPPE 416
 QY 418 SLPLSLVFQNLQVIRGRILHNGAVS-LTLQGLISWGLRSLRELGSGLALIHNTLCLF 476

```

Db 417 NMISLSTFQNLLEIRGTTTSRGSSFTVVOYRHLQWGLSLKLEVSAGNVLKKTIDLARY 476
Qy 477 VHTVPMQDLFRNPHQALHTNANPEDECVGBGLACHQLARGHGWGFPQCVNGSOFILR 536
Db 477 ANTIMMRRLFRSEQSIIEYDART-----ENQCNNECEDGEGWGSPMTVCVSLHVD 529
Qy 537 GQELVEBCRYLQGLPREYVNAARHCLPCHPECOPOGNSVTFEGPEADCCVACAHYKDPFC 596
Db 530 GGRCVASCNLQGGPREAOVDGRCVQCHQECVOTDSTLTCTGEPANCSAHAFQDQPOC 589
Qy 597 VARCP---FNNFTVSFWLRVPKXVSASHLEPDEBACQPCPINTHSCVDDDDXGCPAQR 653
Db 590 IPRCPHGLDGDTLNWKYA-----DKMQCCQPCQNCQTCQSGSGPLSGCRD-1 638
Qy 654 ASPLTYSANVGLLVVGVGVGCLILKRRQCKIRKTMARLLQELVEPLTPSGAMP 713
Db 639 VSHSLAVGLVSGLLITVALLIVLLRRRIRK-RRRTTRCLQEXELVEPLTPSGQAP 697
Qy 714 NQAGRLKETELKRVKVLGSGAFGYKGIWIPDGENVAKPVALKYLRENTSPKANKET 773
Db 698 NQAFRLIKETEFKKDRLSGAGTYKGLMNDGNNIRIPVALKYLRENTSPKANKQEV 757
Qy 774 LDEAYVWAGVSPVSRLLGICLTSTYQVLTQMLPQYGLDHRNNGRLGSDDLLWCM 833
Db 758 LDEAYVWASVDHPVCRLLGICLSAVQLTQMLPQYGLDLYVQHGERICQGLWLMVCV 817
Qy 834 QIAKMSYLEDVRLVHRDLAARNLVKSPNVKITTDLGLRLDIDETEVHADGKVPK 893
Db 818 QIAKMSYLEDVRLVHRDLAARNLVKSPNVKITTDLGLRLDIDETEVHADGKVPK 877
Qy 894 WMALESILRRRTHQSDVSYGVTVWELMTFGAKPYDQIPAREIPDLEKGERLPQPPIC 953
Db 878 WMALESILQMTYTHQSDVSYGVTVWELMTFGSKPYDQIPAREIPAKELASVLENGERLPQPPIC 937
Qy 954 TIDVYMMVKKWMDSECRPRFRELVSFSMARDPQRFVYIQNEDIGPPASPLDSTYRS 1013
Db 938 TIEVYMTILKCMWMDPSRPRFRELVSFSQWADPSRYLTQ---NLPSLSRRILFSR 994
Qy 1014 LLEDDMDGLVDAEYLVPOQGFPCPPAPAGAGVHRRHSSSTRSGGDLTLGLEPSE 1073
Db 995 LLSDD--DVYDADBEYLLPYKRI-----NRQGS----- 1020
Qy 1074 EEARSPPLAPSEGGSGDVFDGDLGMGAAGLQSLPTDPSRLQYSDPY-PLPSTDG 1132
Db 1021 -----EPCIPPTG-----PVENSTLTINISDPQNALKXLDG 1055
Qy 1133 YVAPLTCSPQPEYVNOFVPRQP-----FSPRE-----GPLP-AARPAGATLERAKTL 1179
Db 1056 H-----EYVNOFSESRSLSDIYNVYEDLTDGQGPVLSQGEATNPSREYVL 1105
Qy 1180 SPGXGVVGVFAFGAVENPEYLTPOGGAAPQHPAPAFDNLTYWDQDPEFGAP 1239
Db 1106 NTNONS--PLVSSGSMDDPDY---QAG-----YQAAF-----LPQAGAL 1140
Qy 1240 PSTFGKPTAEPEYLG 1256
Db 1141 TNGMFLPAENLEYLG 1157

```

```

OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Ploman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erb3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT. HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; M29366; AAA55790.1; -
DR EMBL; M34309; AAA55979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN KINASE TYR; FALSE NEG.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferase; Tyrosine-protein kinase; Atp-binding; Phosphorylation;
KW

```

Alternative splicing.

FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	1342	RECEPTOR PROTEIN-TYROSINE KINASE ERB-B-3.
FT	DOMAIN	20	643	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	644	664	POTENTIAL.
FT	DOMAIN	665	1342	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	709	966	PROTEIN KINASE
FT	NP BIND	715	723	ATP (BY SIMILARITY).
FT	BINDING	742	742	ATP (BY SIMILARITY).
FT	ACT SITE	834	834	BY SIMILARITY.
FT	DISULFID	186	194	BY SIMILARITY.
FT	DISULFID	190	202	BY SIMILARITY.
FT	DISULFID	210	218	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	227	235	BY SIMILARITY.
FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	552	565	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VASAPLIC	141	183	ELISGVTIEKNDLCMDITDMRDIVRDRPDLVVDNGR SC -> GQFVWVSGLTPOQADQVYLDDDDPRLLTSSASK VPVTLAAV (IN SHORT ISOFORM).

Query Match 34.9%; Score 2387.5; DB 1; Length 1342;
 Best Local Similarity 39.9%; Pred. No. 1.7e-120;
 Matches 527; Conservative 193; Mismatches 459; Indels 141; Gaps 34;

Query Match 34.9%; Score 2387.5; DB 1; Length 1342;
 Best Local Similarity 39.9%; Pred. No. 1.7e-120;
 Matches 527; Conservative 193; Mismatches 459; Indels 141; Gaps 34;

Query 10 GLTLALPPGA--STOVCTGTDMKLLPASPENHMLNHLVGGCVQVGNLETLPT 67
 Db 11 GLFSLARGSEVNSQAVCPETLNGSLVTDGAEQVTLTKLVERCVVNGNEIYLTGH 70
 QY 68 NASISFLQDIQEVGVYLAHNVQVPLQRLVIRSTQLEFEDYVALAVDNGDPLNNTT 127
 Db 71 NADISFLQWIREVGVYLVANNEPSTPLPRLVRVSTQYVDKFAIFW-----LNYNT 125
 QY 128 PYVGASPGRLRELQRLSLTEILKGVLIQNPOLCYDITLMDIFHKNNQALTLIDTN 187
 Db 126 ----NSSHALRQLTLTLTSLGVTIEKNDKLCMDITDMRDIVRDRD---AEIVVXD 178
 QY 188 RSRACHPSPMCKSCRCMGSSSEDCSLTRTVGAGG-ARCKGFLPTDCCHQCCAGCTG 246
 Db 179 NGSSCPCHVCKG-RCMGPGSEDCQLTTLTICAPQNGCHCPGNPNQCHDCAGCGG 237
 QY 247 PKHSDDLCLLHPNHSGLCEHCPALVYNTDTFESMNPBGRYTFGASCTYACPYVLT 306
 Db 238 PQTDFCACHFNDSGACVPRCPQPLVYNTLTFQLEPNPHTKLYQGVGVASCPNHFV-V 296

Query 307 DVSGCTLVCPHNGVETADGTQCEKSKPCARVCYGLQMYIKANSKC--IGITBLE- 363
 Db 297 DQTSVCYACPPDKMEVD-KNGLKCEPCGGLCPKACSGTS-----SGSRFQYDSSNIDG 350
 QY 364 PACCXKIFGSLAPSEFGDDPASPANRPLQPELOVPELTLEITGYLYISANQDLSPLDS 423
 Db 351 FVNCSTKLGLDPLITGLNDPWHKIPALDPEKLVNFRVREITGLVNTQSWPHMNF 410
 QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLISWIGLSRLSELGSLALIHNNHLCFVTPVW 482
 Db 411 VFSNLTITGRSLVYNGFSLIMKNLNTVSLGFRSLKEISAGRIYISANQLCYHSLNW 470
 QY 483 DQLEPRNQHALLHTA-NRPEDECVGSLACHQLCANHCWCPPTQCVNCSQFLRQGEV 541
 Db 471 TKYLRGTEBRDLIKNRRPRDCAVAGKVCDDLSSGGCGPFGCCLSRNYSRGVCV 530
 QY 542 EECRVLYGQPREVYNAHCLPCHEPCQPNQSVTCGPREADQVCAAHYDPPFCARCP 601
 Db 531 THCNPLNGEPRFAHEACSCPCHECPMEGATCMGSGSDTCAQCHFRDGHCVSSCP 590
 QY 602 FNNFTVSPWLVRKVASHLEPDEBACQPCPIINTHSCVYDLDDKCGCPAEQRA----SPL 657
 Db 591 HGVLAGK-----GPYKYPDVQNECPCHENCTQCGKPELOCLQGLTVLIGKTHL 642
 QY 658 TSIVSAVAGLLVYVUGVYFGILKROQKIR-KYMRRLQETELVEPTBGCAMPNQ 716
 Db 643 TVALTYIAG--LVVIFMMLGGTFLVWGRRIQKRAMRRLERGESIEPLDPS-ERANKV 699
 QY 717 QKRILKETELRKVYLSGAGFTVYKGMIPDSENYKIPAVIVLRBENTSFRANKELDE 776
 Db 700 LAIPFETELRKLVLSGSGVFGVHGVMIPESES.KIYVCIVYIDKSGRGSFOAVTDH 759
 QY 777 AYVMAVGSPYVSRLLIGICTSTVQ:VTLQMPYGLLDHYRENRGLSGQDLLNWCQIA 836
 Db 760 MLAGSLDHAHIVRLGLCPGSSGLQVLTQYLPFGSLDLHYRGALGQLLNMGVQIA 819
 QY 837 KGMVYEDVRLVYRDLAARNVLYKSNHXTIDFGARLLDDETYHADGGVPRKQMA 896
 Db 820 KGMVYEEBGMHKNLAARNVLYKSNHXTIDFGARLLDDETYHADGGVPRKQMA 879
 QY 897 LESILRRPTTHSDVNSYGVYVWELMTFGAKPYDGIIPAREIDLEKGRLPQPCITD 956
 Db 880 LESIHGKTHGSDVNSYGVYVWELMTFGAEYAGRLAEVDLEKGRLLQPCITD 939
 QY 957 VYVIMYKCMIMISECPRRRELVSERAPRQPRVYIQNEDLGA---SPUDSTFYRS 1013
 Db 940 VYVWVYKCMIMIDENIRPTKELANETRYARDPPRYLVTKRES-GPGIAPGEPHGLTNK 998
 QY 1014 LLEDDMDGLVDAAEYLVPOQGFPCDPAPAGAGVNHHRSSSTPSGGDLTLGLEP-S 1072
 Db 999 KLEVEVLEPELDLDDLEAED-----NLATITLQSLALSPVGTUN 1039
 QY 1073 EEDAPRSLAPSEAGSDVFDGLGKAAGKLGSLPTHD-PSPLQRYSEDPVPLP--- 1127
 Db 1040 RPRGQSLSLSPSSGY-MPMNQNLGESQESAVSSSEPCRPVSLH-----PMRPGCL 1092
 QY 1128 --SETDGYXA-----PLTSGQPE-----YVNOVPRPQPPRPRGP--- 1162
 Db 1093 ASESSEGHWTGEAEIQEVSMCRSRSRSRSPRPGDSAYHSQRSLTLTPVPLSPGLE 1152
 QY 1163 -----LPAAPGATLERAKTLP-S-GKNGV-----KDVFAFGAVENPEVLTPOG 1207
 Db 1153 EEDVNGVYVWPDTHLKGTSRREGTSLVGLSGLVLTGTEBEDD-----EVEYEMNRR 1204
 QY 1208 GAAPQHPHPPAPSPAFDNLVYD-----QDPPERCAPSTKGTFTANPEVL 1255
 Db 1206 RHSP-PHPRPSSLLELGEVYWDVGSLSLGSLSQSCPLHPVPIPTAGTTPDEDVYM 1263

RESULT 10
 ERB3 RAT
 ID ERB3 RAT
 AC Q62799; Q62955; STANDARD; PRT; 1339 AA.

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (c-erbB3).
 GN ERBB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxId:10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; Tissue=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sterke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RL (2)
 RP REVISIONS TO 85, 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; Tissue=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Fromert P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neurotrophins and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U29339; AAC8498.2; -
 CC EMBL: U52310; AAC83050.1; -
 CC HSSP: P11362; 1FCR.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PR01009; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT FT 643 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 20 643

FT TRANSMEM 644 662 POTENTIAL.
 FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).
 FT 183 259 CYS-RICH.
 FT DOMAIN 707 964 PROTEIN KINASE.
 FT NP_BIND 713 721 ATP (BY SIMILARITY).
 FT BINDING 740 740 ATP (BY SIMILARITY).
 FT ACT_SITE 832 832 BY SIMILARITY.
 FT DISULFID 186 194 BY SIMILARITY.
 FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 210 216 BY SIMILARITY.
 FT DISULFID 214 226 BY SIMILARITY.
 FT DISULFID 227 235 BY SIMILARITY.
 FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC). (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC). (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F402BBDF5LE CRC64;
 Query Match 33.9%; Score 2316.5; DB 1; Length 1339;
 Best Local Similarity 40.4%; Pred. No. 1,1e-116;
 Matches 521; Conservative 166; Mismatches 437; Indels 165; Gaps 35;
 QY 3 LAALCRWGLLLALLPREGA---STVCTGTDKRLRPASEFTLDMYLRHYGGCCVYQGN 59
 DB 7 LQVLC---FLSLARGSEWNGQAVCPGLNGLSTGADNQTLYLKYKCEVYMG 62
 QY 60 LELTYPTNALSFLQDIOEVGVYLAHQVQVFLQRLRIYRGTOLEFDNYALAVLDN 119
 DB 63 LEIVLGHNDLSFLQIRLEVTVGVYLVANNEFVLLPRLRYVRGTQVVDGKRAIVM-- 120
 QY 120 GDPINNTPTVYTGASPGGLRELRSLTEILKGVLLIQRPOLCYQDTILMKDIFKKNOL 179
 DB 121 ---LNYVT---NSSHALRQLKFTQLTEILSGGVYLEKDKLCHEMTIDIRDIVRVR-- 170
 QY 180 ALLLIDNNRBRACHPSPMCKSGRCGSESEDCSLTRYACAGGC-ARCKGPLPTCCHE 238
 DB 171 GAETVKNNGANCPCHVECKG-RKMGPEPDDQILIKTICAQCGRCGPRPNQCCHD 229
 QY 239 QCAAGCTGPRKNSDCLACLNHNSGICELNCPALVTYNTDTFESMPNPEGRYTPGASCYA 298
 DB 230 ECAAGGCGGPGDTCFCACRPNDSGACVPCRPBELVYNNKLTFLQEPHPHTYGVGVAS 289
 QY 299 CPNYVSLTVDSCTVCPRLNNOVLTAEQGRQREKSKPARVCYL--GMQYTKANSKF 356
 DB 290 CPNHFV-VDDTFVPCACPPDKMEVD-KHGLKMEPCGGCLCPKACBETGSGSRQYTVSSN 347
 QY 357 IGTLELFAAGCKKIFGSLAVLPSPFGDPSNAPLOPEQLQVFETLEITGLYISAMP 416
 DB 348 ID----GFVYCTIKLGNLDFLTGLAVVDPHMKIPLADPEKLVNFRIVREITGLNTQSNP 403
 QY 417 DSLPDLVSFQNLQVITGRILHNGAYS-LTLQGLIGISWLGRLSRLDELGSLALIHNTHLIC 475

```

Db 404 PHMNFVFSNLTITGRSLNKGFSLLIMKNLVITLGRFSLLEISAGVYISANQOLC 463
Qy 476 FVHTVMDOLFERNHQAALLHTA-NRPEDECVBEGIALCHOLCARGCHMCPPTOCVNCSCF 534
Db 464 YHHSILNMTLRLRGSEERLDIKYDRPLGECIABEGKVDPLCSSGCGCPGGQCLSRNY 523
Qy 535 LRGECEVEBCEVLOGLREYVYNAHCLPCHPECCOPONGSTYCFEPADQCAAHYKDP 594
Db 524 SREGCVTHCNFLOGEPREFEHAQCSCCHPECLPMEGISTCNGSGSDACARAHFRDGP 583
Qy 595 FCVACRPFNNFTVFWMLRVPKVASHLPEDEGACQCPINCTHSC--VDLDXGCAEQ 652
Db 584 HCVNSCPHGLIGAK-----GPYKYPDAQNCRCRCHENCTOGCGPELODCLGAEV 635
Qy 653 RASPLTIVSAVVGILLVVLGVVFGILIRROOKR-KYTMRLLOETELVPLTPSGA 711
Db 636 LMSKPHLVIAVTVG--LAVILMIGSGFLYWRGRIQNKRAMRYLERGESIEPLDPS-E 692
Qy 712 MPNQAQRIILKETELRKVKYLGSGAGTGVYKGIPIGGENVKI PVATKYLRNTSPRANK 771
Db 693 KANVYLARIFKETELRKVKYLGSGAGTGVYKGIPIGGENVKI PVATKYLRNTSPRANK 752
Qy 772 EILDEAYMAGVSGPVYSRLIGLCTSTVQVLTQVLMYGGCLDHYRENRGLSGODLLNM 831
Db 753 AVTDHMLAVGSLDHAHIVRLIGLCPGSSLCQVLTQVLPGLSLDHVKQHRETLGPOLLNM 812
Qy 832 CMQAKGMSYLEDVRLVHRDLAANVYKSPNHYKITDPEGLALDLIDETRYHADGGKVP 891
Db 813 GVOIAKGMYYLEESMHWRLALRNWMLKSPSOQVADLFVADLLPDDQQLHSAEAKTP 872
Qy 892 IKMMALESILRRRTTHSDVSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPOPP 951
Db 873 IKMMALESILRRRTTHSDVSYGVTVWELMTFGAEYAGLRLAELPDLLEKGERLPOPP 932
Qy 952 ICTIDVYMWKCMKIDSECRPPREFVSEFSRMDARPOFVVIQNEIDLPASPLDSTFY 1011
Db 933 ICTIDVYMWKCMKIDSECRPPREFVSEFSRMDARPOFVVIQNEIDLPASPLDSTFY 989
Qy 1012 RSLLEDDMDGLVDAAEELVPOQGFPCPDAPAGAGVNHRRHSSSRSGGGLTGLFE 1071
Db 990 PSVUTTEL-----QEAELEPEL-----DLDDLEA 1015
Qy 1072 SEE-----EAPRSLAPSEG-----AGSDVFPDGLGMAAKG 1103
Db 1016 EEEGLATSLGASLPTGTLTTPRGSSQLSPSSGVPMWQSSLGACLDASVLAGREQF 1075
Qy 1104 LQSLPTHPDPEPLQRYSDPTVPLPSETDGVY----APL-----TC-----SPOE--- 1144
Db 1076 SRPISLH-PIPRGR-----PASESEGHVTSSEAELOEKVSVCRSRSRSPRPGDS 1127
Qy 1145 -YVNOPDVPRQPPSPREGP-----LPAPAPAGATLERAKTISP-GKNQGVY----- 1187
Db 1128 AYSQSRSLTPTPLSPGLEBEDNGVYMPDTHLRGASSSEGLTSSVGLSVLTGTEB 1187
Qy 1188 --KDVAFAGAVENPEYLTPOGGAAPPP 1216
Db 1188 EDED-----EYEVNKRKRGSP-PRPP 1209

```

```

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
RX NCBI_TaxID=7227;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schupbach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN (12)
RP REVISIONS.
RA Clifford R., Schupbach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
RN (4)
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
RN (5)
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
RN (6)
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Abrita J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Bens P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Daverport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwan C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spadling A.C., Stapleton M., Stump R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,

```

RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [7]
 RP SEQUENCE OF 959-1078 FROM N.A.
 RC STRAIN=Daekwanyeong;
 RX MEDLINE=85137938; PubMed=2983232;
 RA Wadsworth S.C., Vincent W.S. III, Bilocheau-Wentworth D.;
 RT "A *Drosophila* genomic sequence with homology to human epidermal
 RT growth factor receptor.";
 RL Nature 314:178-180(1995).
 RN [8]
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=92038942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.;
 RT "Interallelic complementation among DER/fib alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW.
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.;
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*
 RT EGF receptor.";
 RL Cell 89:13-16(1997).
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTRAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF CENTRAL CELL FATES. MAINTENANCE OF AMNIOSEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III, ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF052754; AAC08536.1; -
 DR EMBL: AF052753; AAC08536.1; JOINED.
 DR EMBL: AF052754; AAC08535.1; -
 DR EMBL: AF052752; AAC08535.1; JOINED.
 DR EMBL: K03054; AAAS1462.1; -
 DR EMBL: K03417; AAAS1460.1; -
 DR EMBL: K03416; AAAS0965.1; -
 DR EMBL: K03418; AAAS1461.1; -
 DR EMBL: AF109077; AAD26134.1; -
 DR EMBL: AF109078; AAD26132.1; -

DR EMBL: AF109082; AAD26133.1; JOINED.
 DR EMBL: AF109078; AAD26133.1; -
 DR EMBL: AF109084; AAD26133.1; JOINED.
 DR EMBL: AF109079; AAD26130.1; -
 DR EMBL: AF109081; AAD26130.1; JOINED.
 DR EMBL: AF109079; AAD26131.1; -
 DR EMBL: AF109083; AAD26131.1; JOINED.
 DR EMBL: AF109080; AAD26135.1; -
 DR EMBL: AE003454; AAF46732.1; -
 DR EMBL: X02293; CA626157.1; -
 DR EMBL: X78920; CAAS5523.1; -
 DR EMBL: X78918; CAAS5521.1; -
 DR EMBL: X78919; CAAS5522.1; -
 DR PIR: A00640; GOFPE.
 DR HSP: P11352; IFCX.
 DR Flybase: FBgn0003731; Egfr.
 DR Interpro: IPR000494; EGFR_Ldomain.
 DR Interpro: IPR000719; Euk_Pkinase.
 DR Interpro: IPR002174; Furin-like.
 DR Interpro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_Ldomain; 2.
 DR PRINTS: PR001001; TYRKINASE.
 DR Prodom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 7.
 DR SMART: SM00219; TYRKC; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE: PS50011; PROTEIN KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 31 1426 POTENTIAL. EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT NP BIND 938 1198 PROTEIN KINASE.
 FT BINDING 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 28.7%; Score 1960; DB 1; length 1426;
 Best Local Similarity 32.5%; Pred. No. 1,4e-97;
 Matches 467; Conservative 183; Mismatches 431; Indels 354; Gaps 40;
 QY 24 QVCTGDMKLRIPASBETHLDLRLHYGCCVYQGNLETTYPT-NASISFLQDIQEVG 82
 DB 100 KICIGKSKSLSPSKNEHHYRLRDRTNCTYVDGLKLTWLPENLDSFLDNTREVTG 159
 QY 83 YVLIANQYRQVPLRLRIVRGTOF-----EDNYALAVLDNGDEPLNNTPTVYGASPGGL 137
 DB 160 YLISHVDVKVFPPLQILRGRTLFS.SVEEKXALFV-----TYSKM 203
 QY 138 RELQRLSRLLEIKGVLIQORNPQLCYQDTILMKDIFPKKNQLALILIDNRSRACHPSGP 197
 DB 204 YLLEIPDLNDVINGVGFNNYNLCHMRITQMSSEIVNSGTDAVYNYDFAPARECPKHE 263
 QY 198 MKGSKSWESSESDCSLRTVCAGGCA--RKGGLPTCCHEGCAAGGTGRHSDCLAC 255
 DB 264 SCTHG-CWGEGRKNQKFSKLTCSPOCAGRGYGRKPRECHLFCAGGCTGTGTQDCLAC 322
 QY 256 LHPNHSIGICELHPALVTYNTDTFSMNPBGRYTFGASCYACFYNYLSTDVSGCLTVG 315
 DB 323 KNFFDEAVSKBECPPMRKYNPTTYVLETPEBKIVAYGATCVKCEP-GHLLRDNGACVSC 381
 QY 316 PLHNOEVTAEQTCCEKSKPCARVCYGLGMQYIKANSKFIGITEL-----EPAGCKX 369
 DB 382 PDKRMDKGE-----CVPNGCPKTC-----PGVTLNAGNIDSRNCTV 422
 QY 370 IFGLAFLPESFEG--DPAASNTA-----PLQPEQLQVETLEETIGYLYISAWPDSLPD 421


```

Db 423 IDNIRILDTFSGFDVYVANYMTGPRYIPLDEREVEFTVEIKIYLNIEGTHPOFEN 482
Qy 422 LSVFONLQVIRGILHNHAY-STLQGLISWGLSLRSLRGLALIHNTLCPVHTV 480
Db 483 LSVFONLQVIRGILHNHAY-STLQGLISWGLSLRSLRGLALIHNTLCPVHTV 480
Qy 481 PMQPLRNPHQALLHTANPEDECVGEGLAGHOLCARHGWGPGFTQCVNCSQFLRGQEC 540
Db 543 RMPALCKEPEOKWVWNNELRADLCEKXGIIICSDQCNEDGCGWAGTQCLTCKNFENGTIC 602
Qy 541 VESCRVYLGRLREYVYARHCLPCHPEGCPNGSVTCGPRADCCVACAHKXDPFCVARC 600
Db 603 IADCGYISNAKY-FDWRTCKICHPCR-----TCNAGADHCQECVHVRGQHCVSEC 654
Qy 601 PFNNF-----TVS----- 608
Db 655 PKKYNDRGVYCRECHATCDGCTGPKDTIGIAGCTCNLAINNDATYKRLKDKRCPDG 714
Qy 609 -FWLRVKSASHLEPDEGACDP-----CPT-----NCTH----- 638
Db 715 YFW-----EVNHPQEGSLKPLAGRAVCRKCHPLCELTNYGVHEQVCSKCTHYKRR 766
Qy 639 -----SC-----VDLDDG----- 647
Db 767 EGCEBGRADHYTDEBERECFORHPEGNGCTGPGADCKSCRNPKLFDANETGPVNSTM 826
Qy 648 -----CPAEOR-----ASPLTS-----IVSAVVGILLVVLG 674
Db 827 FNCTSKCPLEMRHNVYOTAIGPYCAASPRSSKITANDVMMIFITIGAVLPTICILC 886
Qy 675 VVFGILKRSQOKIRKXT-MRRLQETELVEPTLTSFGAMPNQAOMRIKETEYRKYVL 732
Db 887 VV-TYICROKQAKKKTVMYALASCESESEPLRBNIGANLCKRIYDABLRKGVVL 944
Qy 733 GSGAGFTVYKGIWIPDGENYKIPVAIKVIRENTSPRANKIIDEAVYAGVSGPYVSRLL 792
Db 945 GMAFGKRVYKGVWVPEGENYKIPVAIKVIRENTSPRANKIIDEAVYAGVSGPYVSRLL 1004
Qy 793 GILLSTVQCLVTLMPGCLLDVHRENRGLSGODLLNMCQIACXSYLEDVRLVHRRL 852
Db 1005 AACVMSQOMLITQMPGCLLDVHRENRGLSGODLLNMCQIACXSYLEDVRLVHRRL 1064
Qy 853 AANVVLKSPNHVYKIDFGLARLLDIDETEHADGKVPKIMMALESILRRFTHOSDVW 912
Db 1065 AANVVLKSPNHVYKIDFGLARLLDIDETEHADGKVPKIMMALESILRRFTHOSDVW 912
Qy 913 SYGVYTWELMTFGAKPDPGIPAREIPDLKESGRLPOPPICITDVYIMVCKMMDSEGR 972
Db 1125 AFGVITWELMTFGAKPDPGIPAREIPDLKESGRLPOPPICITDVYIMVCKMMDSEGR 972
Qy 973 PREREIVSESRMARDPORFVVIQNEIDL--PASPLDSTFYRSLLED--DWGDLVDAE 1027
Db 1185 PTKKQTLTVAFARADGRYLAIPGDKFTLPA-----YTQDEKDLIRKLAFTTDS 1237
Qy 1028 EYLVPOGCFPCDPAPAGGMVHRRHSSSTRSGGDLTLGLEPSESEAP-----RSPL 1081
Db 1238 EALAKPDDYLOPYAAPGPS-----HRTDCT-----DEMPKLIRYCKDPS 1276
Qy 1082 APEEGAGSDVFDG--DLGMAKAGLQSLPTHPSPLQRYSEDPVPLSEDTGYAPRLT 1138
Db 1277 NKMSSTGDDERDSSAREVGNLR-----IDLPEDEDDYLP--T 1314
Qy 1139 CSBQPEYVNGPDVAPPSPREGPPAPARPAGATLBRAKTLSPGKNGVYKDYAFAGAVE 1198
Db 1315 CQGPNNNNNNM-----NENQNNMAVGAAGYM-----DLIGVPSVD 1353
Qy 1199 NPEYL-----TPOGGAAPQH-----PPAPSP-AFNDLYIWD 1230
Db 1354 NPEYLLNAOTLGWGESPTTGTIGIYMGSGPGMEVKVPMGPSEPTSSDHEYND 1408

```

RESULT 12

```

ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=85228222; PubMed=2988784;
RX Nilgen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726 (1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: M10066; AAA48763.1; ALT_INT.
DR PIR: A00643; TVCHLV.
DR PIR: B00643; TVFVLV.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transfaser: Tyrosine-protein kinase; ATP-binding; Oncogene;
DR Glycoprotein; Phosphorylation.
DR DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 25.0%; Score 1708.5; DB 1; Length 634;
Best Local Similarity 51.5%; Pred. No. 1.6e-84;
Matches 369; Conservative 75; Mismatches 136; Indels 137; Gaps 19;
Qy 587 CAHYNDPEFCVARGCF-----NFTVSPFLARVPKYSASLLEPDEBACQPCPTINCHSCVD 642
Db 3 CAHFDGPHCVKACAGVLENDTL-VKTA-----DANAQCQLCPNCTRGCKG 51
Qy 643 LDDKCPAEORASPLTSIVSAV-V-GILVVLGVVFGILKRSQOKIRKXTMRRLQET 701
Db 52 PGLBCP---NGSKTPIAAGVGGILCTLVVVGIGIYLARR-HYVKRTLRRLQERE 107
Qy 702 LVEPLTSGAMPNQAOMILKETELRKYKVLGSAFGIVYVGIWIPDGENYKIPVAIKVL 761

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 seconds

(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 6841
Sequence: 1 MELALCRMGLLLALLPFGA.....TFKPTPTAENPEYLGHDPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL 21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mnc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_ivirus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	6114	89.4	1259	6	O18735
2	3039	45.3	1209	11	O9QX70
3	3077	45.0	1210	11	O9EP98
4	2687	39.3	1165	13	O9YH40
5	2687.5	39.0	1137	13	O9W6F6
6	2259	33.0	1328	13	P79754
7	2026.5	29.6	1433	13	P79754
8	1871	27.3	419	4	O9UK79
9	1739	25.4	367	11	O8R2X1
10	1697.5	24.8	412	4	O8WYV0
11	1679	24.5	729	15	O86712
12	1677	24.5	962	15	O86714
13	1612.5	23.6	962	15	O86895
14	1604	23.4	545	15	O85468
15	1468.5	21.5	655	11	O9WVFS
16	1452.5	21.2	643	11	O9ERV6

17	1273	18.6	1193	5	O9Y1X8	O9Y1X8 ephydactia f
18	1206.5	17.6	1368	5	O23821	O23821 caenothabdi
19	1176	17.2	1717	5	O26566	O26566 schistosoma
20	1126	16.5	527	13	O90836	O90836 gallus gall
21	1001.5	14.6	478	11	O9ESR0	O9ESR0 rattus norv
22	906	13.2	165	4	O14256	O14256 homo sapien
23	901.5	13.2	599	13	O9PSH2	O9PSH2 gallus gall
24	806.5	11.8	346	13	P11776	P11776 xiphophorus
25	783	11.4	176	11	O923V5	O923V5 rattus norv
26	778	11.0	435	5	O8SZM1	O8SZM1 drosophila
27	754.5	11.0	311	13	O99162	O99162 xiphophorus
28	734	10.7	331	4	O9EUD7	O9EUD7 homo sapien
29	731	10.7	1671	5	O9N0V5	O9N0V5 biophalari
30	723	10.6	149	6	O9BG66	O9BG66 oryctolagus
31	720.5	10.5	1362	13	O9PVZ4	O9PVZ4 xenopus lae
32	693	10.1	1418	13	O93457	O93457 scophthalmu
33	685.5	10.0	1368	13	O8UW85	O8UW85 paraliichthy
34	670.5	9.8	1369	13	O8UW86	O8UW86 paraliichthy
35	668.5	9.8	1472	5	O9U5A8	O9U5A8 bombyx mori
36	665	9.7	1358	13	O73798	O73798 xenopus lae
37	653	9.5	1412	13	O8UW84	O8UW84 paraliichthy
38	644.5	9.4	1418	13	O8UW83	O8UW83 paraliichthy
39	633	9.3	1245	13	O9YGH8	O9YGH8 scophthalmu
40	628	9.2	1371	11	O9QVW4	O9QVW4 rattus norv
41	619.5	9.1	2144	5	O9YD94	O9YD94 drosophila
42	592	8.7	987	11	O91YMO	O91YMO mus musculu
43	587.5	8.6	1036	4	O07912	O07912 homo sapien
44	587.5	8.6	1055	11	O54967	O54967 mus musculu
45	587	8.6	935	4	O96135	O96135 homo sapien

ALIGNMENTS

RESULT 1

ID O18735 PRELIMINARY; PRT; 1259 AA.

AC O18735;
DT 01-JAN-1998 (TEMBUREL 05, Created)
DT 01-JAN-1998 (TEMBUREL 05, Last sequence update)
DT 01-JUN-2002 (TEMBUREL 21, Last annotation update)
DE ExB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_Taxid=9615;
RN [1]
RP SEQUENCE FROM N.A.

RA Yokota H.;
RT "cDNA cloning of erbB-2 from canine mammary gland";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -

DR HSSP; P1362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Purih-like.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match 89.4%; Score 6114; DB 6; Length 1259;
 Best Local Similarity 88.7%; Pred. No. 0;
 Matches 1133; Conservative 44; Mismatches 66; Indels 34; Gaps 4;

QY 1 MELALCNGWLLALLPGCASTQVCTGDMKRLPASPETHLDMRLHYQCCQVQVGNL 60
 DB 1 MELAMCNGWLLALLPGCASTQVCTGDMKRLPASPETHLDMRLHYQCCQVQVGNL 60
 QY 61 ELTYLPANSLSPLODIOEVQGVYLIANQVAVLPORLRIYRGTQLEFEDNALAVLNG 120
 DB 61 ELTYLPANSLSPLODIOEVQGVYLIANQVAVLPORLRIYRGTQLEFEDNALAVLNG 120
 QY 121 DPLNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHNQOLA 180
 DB 121 DPLNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHNQOLA 180
 QY 121 DPLNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHNQOLA 180
 DB 121 DPLNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYODTILMKDIFHNQOLA 180
 QY 181 LTLIDTNRBACHPGSPMCKSGRCWSESEDCQSLTRYCAGGACRCGRLPTDCCHQC 240
 DB 181 LTLIDTNRBACHPGSPMCKSGRCWSESEDCQSLTRYCAGGACRCGRLPTDCCHQC 240
 QY 181 LTLIDTNRBACHPGSPMCKSGRCWSESEDCQSLTRYCAGGACRCGRLPTDCCHQC 240
 DB 181 LTLIDTNRBACHPGSPMCKSGRCWSESEDCQSLTRYCAGGACRCGRLPTDCCHQC 240
 QY 241 AAGCTGPKHSDCLAHFNHSGICEIHCPLATVYNTDTFESMNPGRYTFEGASCVTACP 300
 DB 241 AAGCTGPKHSDCLAHFNHSGICEIHCPLATVYNTDTFESMNPGRYTFEGASCVTACP 300
 QY 301 VYLTSDVGSCTLVCPHANOVTADGTCRCCKSPCARVYCYGMQITKANSKFTIGIT 360
 DB 301 VYLTSDVGSCTLVCPHANOVTADGTCRCCKSPCARVYCYGMQITKANSKFTIGIT 360
 QY 301 VYLTSDVGSCTLVCPHANOVTADGTCRCCKSPCARVYCYGMQITKANSKFTIGIT 360
 DB 301 VYLTSDVGSCTLVCPHANOVTADGTCRCCKSPCARVYCYGMQITKANSKFTIGIT 360
 QY 361 ELEFAGCKIFGSLAFLEPSFDGDPASNTAPLOPPOLOVFETLEITGVLYISAMPDLP 420
 DB 361 ELEFAGCKIFGSLAFLEPSFDGDPASNTAPLOPPOLOVFETLEITGVLYISAMPDLP 420
 QY 361 ELEFAGCKIFGSLAFLEPSFDGDPASNTAPLOPPOLOVFETLEITGVLYISAMPDLP 420
 DB 361 ELEFAGCKIFGSLAFLEPSFDGDPASNTAPLOPPOLOVFETLEITGVLYISAMPDLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRLSGSLALIHNTLHLCFVATV 480
 DB 421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRLSGSLALIHNTLHLCFVATV 480
 QY 421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRLSGSLALIHNTLHLCFVATV 480
 DB 421 DLSVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRLSGSLALIHNTLHLCFVATV 480
 QY 481 PMDOLFNRHQAALLTANRPEDECEGEGACHQCLARGCWSPGTQCVNCSQFRRGEC 540
 DB 481 PMDOLFNRHQAALLTANRPEDECEGEGACHQCLARGCWSPGTQCVNCSQFRRGEC 540
 QY 481 PMDOLFNRHQAALLTANRPEDECEGEGACHQCLARGCWSPGTQCVNCSQFRRGEC 540
 DB 481 PMDOLFNRHQAALLTANRPEDECEGEGACHQCLARGCWSPGTQCVNCSQFRRGEC 540
 QY 541 VECEVYLOGLPREYVNAHRLCPCHPECOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 DB 541 VECEVYLOGLPREYVNAHRLCPCHPECOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 QY 541 VECEVYLOGLPREYVNAHRLCPCHPECOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 DB 541 VECEVYLOGLPREYVNAHRLCPCHPECOPONGSVTCFPEADQCVACAHYKDPFCVARC 600
 QY 601 PENNFTVSFWLEVPVSAASHLEP-----DEBACQCPINCHSCVDLDDKCP 649
 DB 601 PENNFTVSFWLEVPVSAASHLEP-----DEBACQCPINCHSCVDLDDKCP 649
 QY 601 PENNFTVSFWLEVPVSAASHLEP-----DEBACQCPINCHSCVDLDDKCP 649
 DB 601 PENNFTVSFWLEVPVSAASHLEP-----DEBACQCPINCHSCVDLDDKCP 649
 QY 650 AEORASPLTSIVSAVGLLVVGLVGVFGILIKRQCKIRKYMRLLOETELVEPLTPS 709
 DB 650 AEORASPLTSIVSAVGLLVVGLVGVFGILIKRQCKIRKYMRLLOETELVEPLTPS 709
 QY 650 AEORASPLTSIVSAVGLLVVGLVGVFGILIKRQCKIRKYMRLLOETELVEPLTPS 709
 DB 650 AEORASPLTSIVSAVGLLVVGLVGVFGILIKRQCKIRKYMRLLOETELVEPLTPS 709
 QY 710 GAMPQOQWRLIKETELKRVKVLGSGAGTGYKGIWIDGENVKIPIVAKYLRENTSFA 769
 DB 710 GAMPQOQWRLIKETELKRVKVLGSGAGTGYKGIWIDGENVKIPIVAKYLRENTSFA 769
 QY 710 GAMPQOQWRLIKETELKRVKVLGSGAGTGYKGIWIDGENVKIPIVAKYLRENTSFA 769
 DB 710 GAMPQOQWRLIKETELKRVKVLGSGAGTGYKGIWIDGENVKIPIVAKYLRENTSFA 769
 QY 770 NKEILDEAVYVAGVSPYVSRLLGICLSTVQLVQLMPYGCILDHVENRGRLGSOGL 829
 DB 770 NKEILDEAVYVAGVSPYVSRLLGICLSTVQLVQLMPYGCILDHVENRGRLGSOGL 829
 QY 770 NKEILDEAVYVAGVSPYVSRLLGICLSTVQLVQLMPYGCILDHVENRGRLGSOGL 829
 DB 770 NKEILDEAVYVAGVSPYVSRLLGICLSTVQLVQLMPYGCILDHVENRGRLGSOGL 829
 QY 830 NWCQOIAKMSYLEVRLVHRDLAARVLYVSPNHVKTDFGLARLLDIDETEVHADGK 889
 DB 830 NWCQOIAKMSYLEVRLVHRDLAARVLYVSPNHVKTDFGLARLLDIDETEVHADGK 889
 QY 830 NWCQOIAKMSYLEVRLVHRDLAARVLYVSPNHVKTDFGLARLLDIDETEVHADGK 889
 DB 830 NWCQOIAKMSYLEVRLVHRDLAARVLYVSPNHVKTDFGLARLLDIDETEVHADGK 889
 QY 890 VPIKMMALLESIRRFTHOSDYVSGVTVWELMTGAPRYGIPARETDLLEGEPLPQ 949
 DB 890 VPIKMMALLESIRRFTHOSDYVSGVTVWELMTGAPRYGIPARETDLLEGEPLPQ 949
 QY 890 VPIKMMALLESIRRFTHOSDYVSGVTVWELMTGAPRYGIPARETDLLEGEPLPQ 949
 DB 890 VPIKMMALLESIRRFTHOSDYVSGVTVWELMTGAPRYGIPARETDLLEGEPLPQ 949
 QY 930 PPICTIDVYIMVWKWMTDSECRPFRELVEFSFMAADPORFVVIQNEDEDGSPAPLDT 1009
 DB 930 PPICTIDVYIMVWKWMTDSECRPFRELVEFSFMAADPORFVVIQNEDEDGSPAPLDT 1009
 QY 930 PPICTIDVYIMVWKWMTDSECRPFRELVEFSFMAADPORFVVIQNEDEDGSPAPLDT 1009
 DB 930 PPICTIDVYIMVWKWMTDSECRPFRELVEFSFMAADPORFVVIQNEDEDGSPAPLDT 1009
 QY 1010 FYRSLLEDDMDGLVDABEYLVPOQGFPCPDPAAGGMVHHRSSSTRSGGDLTLGL 1069
 DB 1010 FYRSLLEDDMDGLVDABEYLVPOQGFPCPDPAAGGMVHHRSSSTRSGGDLTLGL 1069

DB 1003 FYRSLLEDDMDGLVDABEYLVPOQGFPCPDPAAGGMVHHRSSSTRSGGDLTLGL 1062
 QY 1070 EPSEEAERSPPLAPBSGAGSDVFDGDIQWGAAGLQSLPHDPSPLORYSEDPVLPSE 1129
 DB 1063 EPSEEAERSPPLAPBSGAGSDVFDGDIQWGAAGLQSLPHDPSPLORYSEDPVLPSE 1122
 QY 1130 TDGVVAPLTCSPQPEYVNOQDVPPOPSPREGPLPARPAGATLER-----AKTSPGN 1184
 DB 1123 TDGVVAPLTCSPQPEYVNOQDVPPOPSPREGPLPARPAGATLER-----AKTSPGN 1182
 QY 1185 GVXNDVAFAGAVENPEYLTPOGGAAPQPPAPAFSPAFDNLVYWDOPPERGAPSTFX 1244
 DB 1183 GVXNDVAFAGAVENPEYLTPOGGAAPQPPAPAFSPAFDNLVYWDOPPERGAPSTFX 1242
 QY 1245 GTPTAENPEYLGIDVPV 1261
 DB 1243 GTPTAENPEYLGIDVPV 1259

RESULT 2

Q9QX70 PRELIMINARY; PRT; 1209 AA.

AC Q9QX70; 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=30236888; PubMed=2342466;
 RA Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;
 RA Petch L.A.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 encoded by an alternatively spliced transcript in normal rat tissue."
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N. A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N. A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K.; Dawson T.L.; Baur H.S.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; M37394; AAF14008.1; .
 DR HSSP; P11362; IFCK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Dkinase.
 DR InterPro: IPR002174; Euk_Dkinase.
 DR InterPro: IPR01245; Tyr_kinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recept_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FV; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KM APP-binding; Receptor; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE7FEC1B7773 CRC64;

Query Match 45.3%; Score 3099; DB 11; Length 1209;
 Best Local Similarity 49.9%; Pred. No. 1; 7e-225;

Matches 643; Conservative 161; Mismatches 353; Indels 132; Gaps 31;

QY 3 LAALCRWGLLALLPPGA-ASTQYCTGTDMKRLPASPETHLMLRLHYOCQVOGNLE 61
 Db 15 LAALCAAG-----GALBEKKVCGSTNRLTGLTFEDHPLSLQMFNNCEVVLGNLE 66
 QY 62 LTLPLTNASISLQDIOEVGYVLIANQVRQVPLQRLRVTRGTOLFEDNYALAVDNGD 121
 Db 67 ITTVORNYDLSPLKTIQEVAGYVLIANTVERIPLNLTIRRNALYENTYALAVLSN- 124
 QY 122 PLNNTPTVTGASPGGLRELQRLSTLTELKGVLIQRPOLCYODTILMKDIFKKNQAL 181
 Db 125 -----YGNNTKGLRELPMRNLOELLIGAVRFSNNPILCNMTIOMRTIV-QDPLSN 175
 QY 182 TLIDTNRS-RACHPCSPCKGSGRCWSESSDCCSLTRTVAGGCA-RCKGRLPTDCHEQ 239
 Db 176 MSMDVQRHLTGCPKDCSPNGSCWGEENCKLTKILCAOCCSRRCRGRSPDCCHNQ 235
 QY 240 CAAGCTGPKASDCLACILFHNHSGICELHCPALVTYNTDPFESMPNEGRTFGASCVTAC 239
 Db 236 CAAGCTGPRSDCLVCHRFDEATCKTCCPLMLYNFTTYQMNVNBEKTSFGATCVKRC 295
 QY 300 PNYVLTDSVSCTLVCPHNOEYTAEDGTGRCEKSKPCARVCYGLGMOYIKANSKFIGI 359
 Db 296 PRNVVYTDHSGCVARACGPDYEV-EEDGVSKCKKCDOPCKKVCNG:GIGEFK-DLSTINA 353
 QY 360 TELE-FAGCKKIFGSLAFLEPSFDGPASTAPLOEQLOVFTTEBITGYLYISAMPDS 418
 Db 354 TTKHFFKCTAIGSDHLFLVARGSFTFTPLDRELILTVLVEITLFIQAMPEN 413
 QY 419 LPDLVFNQLVIRGRILHNGAVSLTLOGISMLGRSLRELGSGLALHNTHLCEVH 478
 Db 414 WTLDAFENELIRGRKCHGQFSLAVVGLNTSLGRSLKESIDSDVILISGRNLCYAN 473
 QY 479 TVWMDQLFRNPHOALHTANRPEDECVGSLAHOICANHGCKGPPPTCCVNCGFLRQ 538
 Db 474 TIMMKLFGTPNOKTKIMNRAKCKATNVCNPLCSSESGCEGPEPTCVCQNSRGR 533
 QY 539 ECVCEGRVLQGLFREYVNAHNCPCHECOPONGSVTCSPREDDCCVCAHKKDPFCA 598
 Db 534 ECVCKNILEGEPREVENECICQCHPECLPQMTNITTCGRGDNCTICAHVYDGHCK 593
 QY 599 RCFP-----NFTVSWLVRVYKVASHLPEDEGACQPCPINTHSCVDLDDKGCFAEOR 653
 Db 594 TCSGIMGENNTLV-W-----KFA-----DANNVCHLCHANCTGCABPGKGC--QQR 639
 QY 654 ASP-LTSIYSAVVGILLVVLGVFGI-LIKRQOKIRKTYMRELLQETLYVEPLTPGA 711
 Db 640 EGGKIPSIATGIVGLFTIV-VALGIGLFMRRLQVRRKTLRLQLQERELEVEPLTPGE 698
 QY 712 MPNQAQMRILKETELERKVLGSGAFYVKGIMIDGENVKIPVAIKYLRENTSFKANK 771
 Db 699 APQQAHLRIKETEFKXIKVLSGAFGTYKGLMIEGKVKIPVAIKYLRENTSFKANK 758
 QY 772 EILDEAYVNAVGSPYVSRLLGLCTSTVQLVQLMPCYCLLDHVENRGRLGQDLNW 831
 Db 759 EILDEAYVNAVDNPHVCRLLGLCTSTVQLVQLMPCYCLLDVYEHKDNIGSYQLNW 818
 QY 832 CMOIAGKMSYLEDVRLVHRDLAARNVLYKSPNVKLTDDGLARLLDIDTEYHAGKRP 891
 Db 819 CVOIAGKMYLEDRRLVHRDLAARNVLYKTPQVKTDDGLARLLCAEKEKHAGKRP 878
 QY 892 IKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEXGERLPOP 951
 Db 879 IKMALESILHRYTHQSDVMSYGVTVWELMTFGSKPYDIGIPASEISLLEGERLPOP 938
 QY 952 ICTIDVYMTMVKCMIDSECPREFRELYSEFSRMARDPRFVYIQ-NEULGASPLDSTF 1010
 Db 939 ICTIDVYMTMVKCMIDADSREKREFRILIEFSKMDPRYLVIGODERMLPSPDTSNF 998
 QY 1011 YRSLLEDDMDGLVDAAEYLVQGGFCFDPAPGAGVHHHRSSSTFGSGGDTLIGLE 1070
 Db 999 YRALMEEDMEDVVDADVYLTPQGGF-----NSPST----- 1030

QY 1071 PSEEEAPRSPLAPSEGAGSDVFDGLGMAKGLQSLPTHDPPLQRYSEDETPVLPSET 1130
 Db 1031 -----SRFPLSLISANSN-----SSTVACIRNNSCKVKNDAFQRISSDPTSVLTEN 1080
 QY 1131 --DGVAPLPTCSPQPEYVNOQVDPQPSPREGPIIPARPAATLIERAKTSPGXGVK 1188
 Db 1081 IDDTFL-----FVPEYINQ-SVPRPAGSVQNPVYHNPPLHP-----APGRDLHYQ 1125
 QY 1189 DVFAFGAGAVENPEYL-TPQCGAAPQHPHPAFSPFNDLYVWDQ-----DP----- 1233
 Db 1126 N--PHSNVANSREYLTAAQ-----PTCLSSGFDSSALMTQKSHQMSLDNPDYQD 1174
 QY 1234 --PERGAPSPFTKGTPTAENPEYLGIDVP 1260
 Db 1175 FEPKPAKPNIGIKG-PTAENAEYLRVAP 1202

RESULT 3
 ID Q9EP98 PRELIMINARY; PRT, 1210 AA.
 AC Q9EP98;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor isoform 1.
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid:10090;
 RN [1]
 RC SEQUENCE FROM N. A.
 RP STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maile N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RC SEQUENCE FROM N. A.
 RP STRAIN=C57BL/6J;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Scheel C., Pearse R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maile N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms.";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275366; AAC28045.1; -;
 DR EMBL; AF275364; AAC28045.1; JOINED.
 DR EMBL; AF275365; AAC28045.1; JOINED.
 DR EMBL; AF275367; AAC24386.1; -;
 DR HSRP; P11362; IFGK.
 DR MGD; MGI:95294; Egfr.
 DR InterPro; IPR000345; Cytochrome heme bind.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_Dkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_Chr_Pkinase.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TyrcKc; 1.

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1_
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; Receptor; Transferase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9D32E18 CRC64;

Query Match 45.0%; Score 3077; DB 11; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 7,8e-224;
 Matches 635; Conservative 160; Mismatches 361; Indels 128; Gaps 27;

QY 11 LLLALLPAGAA-STVCTGTGDKLRLPASPEHLDMRLHLYGCGVQVGNLELYLPTN 68
 Db 14 LLLTALCAAGALAEKKKVCQCGTSNRLTGLTFEDHFLSLQRMVNCVGLNLEITYQRN 73
 QY 69 ASLSFLDIOIEVGVYLIHNVQVQLPRLIRYRGQLFEDNVALVLNGLDPLNNTTP 128
 Db 74 YDLSPFLKIOEVAGIVLALNTVERIPLENQILRGALYENTYALALISN----- 124
 QY 129 VTGASPGGLRELQLSLTEILKGVLLQRNPOLCYODTILMKDI---FHKNQALTLI 184
 Db 125 -YGTNRGLRELPRNLQELIGAVRFSNNPILCNMDTILQWRDIVQVFNMSMDL--- 180
 QY 185 DTNSRACHPCSPMKSGSRGWGSESDQGSITRTVCGGCA-RKGLPLPTDCCHEGCAAG 243
 Db 181 -QSHPSGCPKCDPSCGPGSCWGGEENCQKLTKLICQAQCSHRGRGRSPDCCCHNQCAG 239
 QY 244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTPESMPREGRYTFGASCVTACPYNY 303
 Db 240 CTGPRESDCLVQCQFQDEATCKDCTPPLMLNPTTYQMDVNPBESKVFQATCYCKCRNY 239
 QY 304 LSTVSGCTLVCPILHNOEVTAEQDTCRCKSKCAVCGLGMOYIKANSKFIQTELE 363
 Db 300 VVTHGSCVRAQGDYEV-BEDGIRKCKCKDGCRCVKNIGIGERK-DTLISINATNIK 357
 QY 364 -FAGCKIFGSLAFPSFDDPASNNTAPLOPELOVEFTLEETGLYLSAMPDLSPL 422
 Db 358 HFXYCTAISGLHLPLPAFKDSTRTPLDPRLLEILKTYKELTGLLIQAMDNDTL 417
 QY 423 SVFONLOVIRGRILHNGAYSLTLOGLSIWLGRSLRELSGALIHNTHLGFVHTPW 482
 Db 418 HAFENLEIRGRTRQHQGFLAVVGLNITSGLRSLKXISDGVIIISGNRNLGVANTIN 477
 QY 483 DQLEFRNHQALHIANRPEDECEBGLAQHLCARHGMWGPFCQVNCQGLRGQCEVE 542
 Db 478 KKLFGTRQKTKIMNBAEKDCAVNHVCPNLCSSGCGWGPEDCVSCONVARGRECE 537
 QY 543 ECRVLQGLPREYVNAHCLPCHPECCOPNGSVTCFGEADQCVCAHYKDPPEVCARCP 602
 Db 538 KCNILEGEPREFVENSEICIQHPECLPQAMNITCTGHRGPNCTICAHYIDGPHCVKCPA 557
 QY 603 -----NFTVSFMLRVKVASASHLEPDEBGCQCPINCHSCVLDLDCGCPAEQASPL 657
 Db 598 GIMGENNTLV-WRYA-----DANNVCHLCHANCYCGCPGLQCEVWFSGPKI 645
 QY 658 TSIIASAVGILLVVLGVVFGI-LIKRQCKIRKYTRRLQETELTEVLTSPGAMNQA 716
 Db 646 PSIAITGIVGGLFIV-VALDIGLFMRRRHIVKRTIRRLLOEHELVEPLTSGEAPNQA 704
 QY 717 QMRILKTELKAVYLSSGAFTYKGIWIPDGENVKI PAIAIKYLRENTSPPKANKETLDE 776
 Db 705 HLRIKETEFKKIYVLSGAFGVYKGLWIPBEKVKI PAIKELREATSPKANKETLDE 764
 QY 777 AYWVAGSGPYVSLGLICTSTVQLTOLMPYCCLIDHRENNGRIGSODLIMNCQIA 836
 Db 765 AYWVASVDNPHVCLLGLICTSTVQLTOLMPYCCLIDYREKNDIGSGLINMCQIA 824
 QY 837 KQMSYLEDVRLVHDLAARNVLKSPHVKITDFFGLARLLDIDETEHADGKVPKIKMA 896
 Db 825 KQMYLEDRLVHDLAARNVLKTPQHVKITDFFGLAKLLGAEKEHYHAGCAKVPKIKMA 884
 QY 897 LESTLRRTFQSGVMSYGVTVWELMTFGAKPYDGIARELPDLLEGEELTPQPTCTID 956

Db 885 LESTLRRTFQSGVMSYGVTVWELMTFGSKRKYDGIASDISISLEKGERLPQPICTID 944
 QY 957 YVIMIKKCMWIDSECRPRRELVSFRRAPQPFVYIQ-NEDLGPASPLSTFYRSL 1015
 Db 945 YVIMIKKCMWIDADSPKFRLELITLFFSKXARDPQRYLYIQGERHMLPSPTSNFRALM 1004
 QY 1016 EDDMDGLDAEAEVLPQCGFCPCPDPAQAGMHHRRHSSSTRSGGDLTLGLEFSEE 1075
 Db 1005 DEEDMDVDVADAEYLTPQGF-----NSPST----- 1031
 QY 1076 APPSLAPSEAGSDVFDGLGMAKGLQSLPTDPSFLQYSEDPTVPLPSET--DGY 1133
 Db 1032 -SRTPILSLSATSN-----NSTVACINRSGCRVEDAFLQYSSDPTGATVEDNIDDAF 1086
 QY 1134 VAPLTCGPEEVNPDVPPRPPREGPLPARAGATLEAKTLSPGKGVVDVFAF 1193
 Db 1087 L-----PVPEVYNQ-SVYKPRAGSVQNPVTHNOLHP-----AFGRDHYQV--PH 1129
 QY 1194 GGAVENPEYL-TPQGAAPQPPPAFPAPDNLVYWDQ-----DP-----PER 1236
 Db 1130 SNVGNPEYLNTAQ-----PTCLSGGFSPALMTQKSGHMSLDNPVYQDPFPKE 1180
 QY 1237 GAPPTFKCTPTAENPEYIGLDVP 1260
 Db 1181 TRPNGLFKG-PTAENAEYLTVAP 1203

RESULT 4
 Q9YH40 PRELIMINARY; FRT; 1165 AA.
 AC Q9YH40;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 NX NCBI_Taxid=8086;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations.",
 RL Oncogene 16:1661-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Scharl M.;
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: U53471; AAD0500.2; -
 DR HSSP: P11362; 1FGK.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR001245; Tyf_kinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recept_L_domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR Prodom: PD000001; Euk_kinase; 1.
 DR SMART: SM00261; Fu; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE: P500107; PROTEIN KINASE ATP; 1.
 DR PROSITE: P50011; PROTEIN KINASE DOM; 2.
 DR PROSITE: P500109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; Kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7FEE38D877A7AE CRC64;

Query Match 39.3%; Score 2687; DB 13; Length 1165;
 Best Local Similarity 45.1%; Pred. No. 2,5e-194;
 Matches 579; Conservative 161; Mismatches 385; Indels 160; Gaps 32;

QY 1 MELALCRWGLLALLPPG-AAST---CVCTGTDKMLRLPASPEHMLDLRLHLYGQCOV 55
 4 LELLEL-----LTLILLISGRCCSTDPDRKVCQGTSTNQMTY---LDNHVLYKMKMYGSCNV 56
 QY 56 VQGNLELTYPTMASTFLDLOIEVGYVLIHNOVROPFLQRLVTRGTFEDNYAIA 115
 57 VLENLETTYQENODLSFLOISQIEVGYVLIHNEVSTIPLVRLIRIQNIYEGFTLL 116
 QY 116 VLDNGDPLNNTTPTGASPGGLRELQRLSTELIKGGLIQRNPOLCYODTLMKDIFHK 175
 117 VMSNYOK-NPSSP--DYVGYGLKQLOLSNLTSLSGGVKSHNPLICNVETIMWDIVX 173
 QY 176 NNGLALLTIDTNSRACHPCSPCKSGRCWSESSDQSLRTVCAGG-ARCKGLPTD 234
 174 TSNPTWMLIPHAFERCQCKDPCGVNGSCMAPGPGHCQKFTKLLCAEQCNRRCRGPKPID 233
 QY 235 CCECCAGGCTGPKHSDCLAFHNSGICELHCPALVTYNTDFESMNPGRYFGAS 294
 234 CNEHCAGGCTGRATDCLARDNDGCKDTCPPKYIDVSHQVNNPIKYTEFGA 293
 QY 295 CVTACPYNYLSTDVSGCTVCPHNOEYTAEDGTQRCCKSCPKARVCYGLGM---QYI 350
 294 CVEKCPSNVYVE-GACVRSCKAGMLEVD-ENGKRSCKCDGCVPCVCDIGISLNTI 351
 QY 351 KANSKRTIGTELEFACKCKIFGSLAFPSFGDDPKSNAPLQPELOVFELEETGYL 410
 352 ANNSTNIG---SFSNCTKINGDIILNRSFECDDPKYIGPMDELMWLTIVKEITGYL 407
 QY 411 YISAMPDLPDISVFONLOVIRGRIHNGAYS-LTQGLGISMLGRSLRELSGLALIH 469
 408 VIMMPENNSTLSVFNLEIRGRTTFSGFSFVYVQVSHQMLGSLISKEYSAGNVILK 467
 QY 470 HNTHLCTFTVPMQDLFENPHQALLHTANRPEDECEGELACHQDLCARGCHWPGPTCV 529
 468 NTPOLRYASTINMRILFRSEDOSEYDART-----ENQCNNECEDSGCGPPTCV 520
 QY 530 NCSQFLRGECEVECEGVLGLPREYVNAHRCUPHEPCQPOQSGVTCGPEPADQVCAH 589
 521 SLIHYDRGRCVASCNLLQGEPRCAVDGRCVQCHOELVQTDSTLCYGPAPNCSKCAH 580
 QY 590 YKDPPEFCVACRP---FNNFTVSFWLVRPKYSASHLEPDEGACQCPINICTHSCVDLDDK 646
 581 FQDGFQCI PRCPHGMGLDGLTLMKYA-----DMGQCCQCHQNCCTGGCGGPGIS 630
 QY 647 GCPAQRASPLTISYSAVYGLLVVYGVFGLIKRROQKTRKTYMRLLQETLEVPL 706
 631 GGRGD-IYSHSLAVGLVSGLLITVALIVLLRRRIK-RKRTIRRLQEKELVPL 688
 QY 707 TFSGAMPNOQWRIILKETELRKVYLGSGAFGVYKGIWIPGGEVVKIPIVAILYRENTS 766
 689 TFSGAPNOQAFRIILKETEFKDRVLGSAFCTVYKGLMNPGEVIRIPVAILYRENTS 748
 QY 767 PRANKELDEATVMAVGSPPYSLGLICTSTVOLVTLQMPYGLLDHVENRGLSQ 826
 749 PRVNOEVLDEATVMAVSDHPVCRLLGLICTLSAVQVTLQMPYGLLDHYVQHOGRICQ 808
 QY 827 DLANCMQIAKMSYLEDVRLVHRLDAAENVLYKSPNHYKTTDEFLALLDIDELFYAD 886
 809 WLLNMCVQIAKMNVLSEHRLVHRLDAAENVLYKSPNHYKTTDEFLALLDIDELFYAD 868
 QY 887 GSKVPIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKER 946
 869 GSKVPIKMALESILQWTVTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKELASVLENER 928
 Db

QY 947 LPQPICTIDVYIMVWKCMIDSECRPRFRELVSFSSMADPQRFVIONEDLGPASPL 1006
 929 LPQPICTIEVYMIILKCMIDPSSRPFRELVSFSSMADPQRFVIONEDLGPASPL 985
 QY 1007 DSTFYSLLLEDDMDGLDYDAEYLYPQGFPCPDPAAGGVHRRSSSTRSGGDLT 1066
 986 DRRLSRLSSD--DVYDADEYLL-----RYKRIN-RQGS----- 1018
 QY 1067 LGLPSEEEAPRSPAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSSLQSYSEDPY-P 1125
 1019 -----EPILPENG-----PVRENSIALYISDPTQNA 1046
 QY 1126 LPSETDGYVAPLTCSQPEYVYQPDVREP-----PSPRE-----GPLP-AARPAAT 1172
 1047 LEKDDIGH-----EYVNPQSESTRSLDIYNNYEDLTGQWGVSLSSGEATN 1096
 QY 1173 LERAKTSLPGKNGVYKDFAFAGAVENREYLTPOGGAAPQHPPAPSPADNLYWQD 1232
 1097 FSRPEYLTNQNLSL--PLVSSGSMDDEY---QAG-----YQAAF----- 1131
 QY 1233 PPERGAPSTFKGPTAEENPEYLG 1257
 1132 LPQIGALTNQGMFLPAAENLEYLGI 1156
 Db

RESULT 5

ID Q9W6F6 PRELIMINARY; PRT; 1137 AA.

AC Q9W6F6;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 CN BRB4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN.
 RX MEDLINE=99263203; PubMed=10328884;
 FA Dixon M., Lumsden A.;
 RT "Distribution of neuregulin-1 (nr1) and erb4 transcripts in
 embryonic chick hindbrain."
 RL Mol. Cell. Neurosci. 13:237-258 (1999).
 DR EMBL: AF121963; AAD31764.1; -;
 DR HSSP: P11362; IFGX
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Euk_Pkinase.
 DR InterPro: IPR001368; TNFR_C6.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; Tyr motif.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF0068; Pkinase; 1.
 DR Pfam: PF01030; Recep_L domain; 1.
 DR Pfam: PF02757; YLP; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR PRODOM: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00261; Fv; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: P500107; PROTEIN KINASE ATP; 1.
 DR PROSITE: P50011; PROTEIN KINASE DOM; 1.
 DR PROSITE: P500109; PROTEIN KINASE TYR; 1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match

39.0%; Score 2667.5; DB 13; Length 1137;

Best Local Similarity 46.2%; Pred. No. 7.2e-193;
Matches 532; Conservative 171; Mismatches 353; Indels 95; Gaps 27;

QY 161 LCVQDTILMKDIFHKNNQALATLIDTNRSAHPGSPMCKSGRSGESSEDCQSLTRIVC 220
DB 3 LCFADITIMQDIDVRNFMASNFTLVPTNGSSGGRCHKCTG-RCMGPTENHCOITLTKIVC 61
QY 221 AGGC-ARCGPLPTDCCHEGCAAGCTGPGSDCLCLHNSGICELHCPALVTNTDTF 279
DB 62 AQCDCRCYGPVSDCCHEGCAAGCTGPGSDCLCLHNSGICELHCPALVTNTDTF 121
QY 280 ESMRPEGRYTFGASCTVACPNYVSTDVGSCTLVCPHNOETADGTQRCCKSKPCA 339
DB 122 QLEHNNHAKYTYGAFVCKKCPNHFV-VDSSSCVRACPSSKMFV-EENGKMKCPCTDICI 179
QY 340 RUCYGLGMOYIKANSKFIGITILE--PAGCKXIFGSLAPLPEFPDDPSANTAPLOPEQL 398
DB 180 KACDGIIGTSL-VSAQTVDSSNIDKINCTKINGNIFLVTHGDPHTTAINPEKLN 238
QY 399 VETLEITGYLYISAMPDLPDLSVFNQLOVIRGILHNGAYSLLTQGLGISMGLRSL 458
DB 239 IFQTVREITGYLNIQSPENMTDFRVSNLVTIGRALYSGLSLILKQOGITSLQFQSL 298
QY 459 REIGSGALIHNTLHCFVHTVPMQDLFNPQALLHTANREDECVGEGLAGHQLCARG 518
DB 299 KQISAGNIYITDNSNLCYHNTVMTSLFSTPQKTIVHKKKAENCTADGVVNCBLSSD 358
QY 519 HGWPGPTQCVNCSQPLRGQECVEECRVQGLPREVYNAHCLPCHPEQCP-ONCSVTCF 577
DB 359 GCMGPRPDCLCKRIRIRTCIESNLYDGEFREPANOSVMECDPOCKEMDMITCY 418
QY 578 GPRADQVACAHYKDPFCVACP-----FNNFVYSFWLRVKVASHLEDEBACQPC 632
DB 419 GPRPDCTCFHFKDPPNCEKPDQLOGANSFIFY-----ADEDECHPC 465
QY 633 PINTCHSCVLDLDDKGC-----PAEQASPLTISIVAV-GILLVVLGVVFGIL 680
DB 466 HPRCTGCGRASHDCIYPTWRQSTLPQAR-TPL-IAAGVIGGLFIVIMGLTAY 522
QY 681 IKRQOKIRKYTRMLLOETELVEPLTPSGAMPNQAOMELKTELRYKVLGSAFGTV 740
DB 523 VRRKSIR-KKRALRRL-ETELVEPLTPSGTAPNOQALITKELTKRYVLSGSAFGTV 580
QY 741 YKGIWTPDEGNKXIPAIKIVLRBNTPSKANKELIDBAVYAGVSPYSLIGLITSLV 800
DB 581 YKGIWVBEGETYKIPAIKILNETTGPKANVEFMDALIMASMDHPLVRLGVCLSPI 640
QY 801 QLVTLMPYGLLDHVRNRSGLSGODLNMCMQIAKMSYLEVALVHRDLAARNVLK 860
DB 641 QLVTLMPHGCCLLDYHEKNDNGSGLLNMVCQIAKGMWYLEBRLLVARDLAARNVLK 700
QY 861 SPNHVKITDFGLARLLDIDETEHADGKVPIKMALESILRRFTHOSDWSYGVYWE 920
DB 701 SPNHVKITDFGLARLLDIDETEHADGKVPIKMALESILRRFTHOSDWSYGVYWE 760
QY 921 LMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYIMYKCMWIDECPPRELV 980
DB 761 LMTFGKPYDGIPTREIPDLEKGERLPQPICTIDVYIMYKCMWIDECPPRELV 820
QY 981 EFSRMARDPQREYVIONED-LGPASPLSTFYSRLJEDDDMDGLVDAEYVLPQGFPC 1039
DB 821 EFSRMARDPQREYVIONED-LGPASPLSTFYSRLJEDDDMDGLVDAEYVLP 879
QY 1040 DPAFGAGVYHHRHRSSTRSGGDLTLGLEPSEBEPARS--PLAP-SEGASDVVDGL 1096
DB 880 PPIYSTRIRIDSRNRFVTRDGYAAEQV-PMYTRAPGCTIPERAVAGAGATAEI 938
QY 1097 GNGAAGLQSLPTHDSPLOQYSEDTVPPLPS-----ETDGYAFLTCSQPEYVNP 1149
DB 939 CNGTLKQVATLAKEDSTQRYSDPTVFIPEVINGLEDDEGYTMPMDKXTDYLNV 998
QY 1150 DVTRPPSPRREGFLPAA-RPAGATLERATKLSGKGVYKDYF-----AFGAVENP 1200

DB 999 EENPFVRRKXKQDQAVDNDEYHN-----ABNGOPKADEYVNEPLYINTEANTLENA 1051
QY 1201 EYLTPOGGAAPQHPHPPAFSPARDNLYMDQDPPEBGA--PSTFPKGTPI----- 1248
DB 1052 EYL-----KNLPEKAKKAFDNDPYMNSLPSTLQHPYLDSEYKPYKONGRI 1103
QY 1249 -----AENPEYL 1255
DB 1104 RPIVAENPEYL 1114

RESULT 6
P79754
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Erib3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.
RL Genome Res. 9:251-258 (1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFG.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Dkinase.
DR InterPro; IPR002174; Eurin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Dkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_Dkinase; 1.
DR SMART; SM00261; Fu_3.
DR SMART; SM00219; TykC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ
SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;

Query Match 33.0%; Score 2259; DB 13; Length 1328;
Best Local Similarity 39.7%; Pred. No. 7.7e-162;
Matches 514; Conservative 157; Mismatches 422; Indels 202; Gaps 33;

QY 9 WGLLALLP--GAATQ---VCTGDMKRLTPASPEHLDMLRHLYQCCQVYQNL 62
DB 4 WRLTLCVASSLRASSQGTQEAQVCGQNLSTSGSENGYNLNKDKYKCEIIMGL 63
QY 63 TYLPNLSLFDIIGVQGYVLIANQVQVPLQRLIRYRGQLFEDNATALAVDNDP 122
DB 64 TQISNNDPFLTKIRIVETGVLIAMNHPOEIPLGQLRVIRGNSLYERRPALSF 120
QY 123 LNNTPPTGASPGGLREQLRSLTEILKGVLIQORNLQCYOPIILKDI 182
DB 121 ----YPDG--PSGLNQLGLMNLTEILDGGVQIINKYLYGPRVYTRDI--RNDAPIE 173
QY 183 LIDTNRSAHPGSPMCKSGRCWGESSEDCQSLTRIVCAGC-ARCGPLPTDCCHECA 241
DB 174 IQFNGERGVCV---KSC-GNYCWGPRKQDQIILTKIVCAQCDRCFGTSPROCHTECA 229
QY 242 AGCGPRGHSCLALHNSGICELHCPALVTNTDFFESMPREGYTFGASCTVACPY 301
DB 230 AGCGPLDPTDCAFRLPNDGACVPCPQTLINNKOTFQMETPNNAKYQYGSICVSQCP 289

QY 302 NYLSTDVSGCTLVCPHNOVTAEDGTOR-CEKSKPCARVCYGLGMOYITANSKFIGIT 360
 DB 290 HFV-VDGSSCVSVCPDPMVEV--EKGSGQCELSGCLCPKCEGTGAE----QRYDSS 342
 QY 361 ELE-FAGCKKIFGSLAFPESEFDDPASNTAPLOPEQLQVETLEETIYGLYISAMPDSL 419
 DB 343 NIDSEINCTKOGSHFLVLTGLDDFKNVPPPLAKKLEVRITREITDILNITSWREL 402
 QY 420 PDLSPQNLQVIRGRILHNGAVSLTQGLISWGLSRSLRELGLALIHNTLCEVHT 479
 DB 403 NDLVSFSLTTIQGSLFKRPSLMWMRIPTLTSLGRSLREISGSVYISQNALCYHHT 462
 QY 460 VPMOQLEPNPH-QALHTANRPEDECEGGLAQOLCARGCMWBPQPCNCGOFLRGQ 538
 DB 463 VNMTOLEFGRKRVANSLNSRPMACVADGRVCPDSCSGCMWGPQCCSLKNYSIHG 522
 QY 539 ECVEECVYLGLEPREYVNAH-CLPCHCECPONGSVTCFGEPAQOCVACAAYDPPFCV 597
 DB 523 TCVAGCHFNHSGIPREFAGLNGVACVCHPECKPQTKASCTGPADECACTKPRDGPYCM 582
 QY 598 ARCP--FNNFTVSFWLRVPKVSASHLEPDEGAQOCPINCTHSCVLLDDKGCABEORAS 655
 DB 583 SSCPAGVNDGEXGLIFK-----PNREGHEPCQONCTQSGSGGLNDG---LEAA 630
 QY 656 PLTSIVSAVGLLVVLGVVF-----GILIKRQCKIRKYMRLQCTELVEPLTP 708
 DB 631 RLTISSGQITGALGVPAGLIFCLVLPFLGMLYHRLGALIRKRAMRRLLESGEFEPDGP 690
 QY 709 SGAMPNQAKMRLKTELEKRYKVLGSAFGVTVYKIMIPDGENVKIPVALVLENTSPK 768
 DB 691 -GEGKTGHARILKPSDLRKIKPLGSGVGFVGSXGFWIPBEETVKIPVALKITIDSSGRQ 749
 QY 769 ANKEIDEAYVWAGSPYVSRLLGICLSTVOVLTOMPRGCLLDHRENRGLSGODL 828
 DB 750 TFEITDULMSGLDHPYIVRLGICGTCLOLVLTQSSGSLSEHROKTKISLDPQL 809
 QY 829 LNMCMQIAKMSYZEDVLRHDLAARNVAVKSPNHVKITDFGLARLIDDETYHADGG 888
 DB 810 LNMVQIAKGMVYLEEHVYHKNLAARNILKNDYQVQISDYVADLLYPDDKKYVYSET 869
 QY 889 KVPITKMALESILRRFTHOSDWSYGTWELTFGAKPFDGIPAREIPLEKGERLP 948
 DB 870 KPIKMALESILFRYTHOSDWSYGTWEMMSFGAEPYASVQPEVPSVLEKGERLS 929
 QY 949 QPCTIDIVYIMVCMWIDSECRPRELSEFSRMAPDQRFVVIQNEIDLPASPLDS 1008
 DB 930 QPACTIDIVYIMVCMWIDSECRPRELSEFSRMAPDQRFVVIQNEIDLPASPLDS 980
 QY 1009 TFRSLLEDDMDGLVDAEYLVPOQGFCCPDPAAGAGVHHRHSSSTRSGGDLTLG 1068
 DB 981 -----EDSGMGERL-----RGSER---GLEAD 1001
 QY 1069 LEPSFEAPRSPPLAPSEGAGSDVFDGLGNG--AAKGLOSLPTHDSPLD----- 1116
 DB 1002 LEEDEE-----GLSDRAFTSLQPSFWSWSPSINSYVMWTQ 1040
 QY 1117 -RYSEDPVPLPSETDGYVALTCSPOP-EVYNQ-----PVYRPPSPPRE 1160
 DB 1041 LRYD-----FAVQGGHIGYLPMSPSVDITRQLWYQSRSLSVRTLPRDSARBSRE 1094
 QY 1161 GPL--PAAPAGALLERAKTILSPKNGVYKVFAGAVENPEVILTQGGAAAPPHPPA 1218
 DB 1095 AELCEDGAQAGIRVR-----FGSERGN-----POGG----- 1122
 QY 1219 FSPAFDNLVYWDQDPPERGAPSTFGKTPTAENPE 1253
 DB 1123 -----QORKLSTASSPSFXTMAADEDE 1146

RESULT 7
 Q9BIH9
 ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
 AC Q9BIH9;

DT 01-JUN-2001 (TEMBUREL 17, Created)
 DT 01-JUN-2001 (TEMBUREL 17, Last sequence update)
 DT 01-JUN-2002 (TEMBUREL 21, Last annotation update)
 DE Putative epidermal growth factor receptor (Fragment).
 GN EGR.
 OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 OC Anopheles.
 NC NCBI_Taxid=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SU4;
 RA Lycett G.J.;
 RT "Cloning, expression and localisation of the Anopheles gambiae
 RL epidermal growth factor receptor."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ31655; CAC35008.1;
 DR HSSP; P11362; 1PGX
 DR Interpro; IPR000345; Cyto_heme_bind.
 DR Interpro; IPR000494; EGR_L_domain.
 DR Interpro; IPR000719; Euk_kinase.
 DR Interpro; IPR002174; Furin-like.
 DR Interpro; IPR002290; Ser_thr_kinase.
 DR Interpro; IPR001245; Tyr_kinase.
 DR Pfam; PF00757; Furin-like.1.
 DR Pfam; PF00669; kinase.1.
 DR Pfam; PF01030; Recep_L_domain.2.
 DR PRINTS; PR0109; TYRKINASE.
 DR Prodom; PD000001; Euk_kinase.1.
 DR SMART; SM00261; Fuf_7.
 DR SMART; SM00220; S_TKc.1.
 DR SMART; SM00219; TYKc.1.
 DR PROSITE; PS00180; CYTOCHROME_C; UNKNOWN_4.
 DR PROSITE; PS00107; PROTEIN_KINASE_APF.1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM.1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR.1.
 KW Receptor.
 FT NON TER.
 SQ SEQUENCE 1433 AA; 159585 MW; 83D9D88967724F07 CRC64;
 Query Match 29.6%; Score 2026.5; DB 5; Length 1433;
 Best Local Similarity 32.8%; Pred. No. 3.5e-144;
 Matches 474; Conservative 199; Mismatches 395; Indels 379; Gaps 39;
 QY 26 CTCTDKKRLPASPETHLMRLHYGCGVYQGNLELTYLPNNAISLFIQIDQEVGYVL 85
 DB 1 CIGTNGMSVPAAREHYKRLDRYNTCTVYGNLEITWINTDINFLQHIREVYGYVL 60
 QY 86 IAHNOVROYPLQRLRYRGOLF-----EDNYALAVLDNGDPLNNTPTVGASPGGLREL 140
 DB 61 ISLYDLPQVILPRLQIIRGRITTFKLKMEAEAGLFV-----SFSMNTL 104
 QY 141 QRSLEILKGVLIQIRNPOLCYQDITLWKDI-FKRNQALATLIDTNSRACHPCSPMC 199
 DB 105 ELPALRDIIGSGVGFNNYNLGHKMSINNEIILAPQTSMTQTFNFSSEPRVCPCHPSC 164
 QY 200 KSRRCWGESSEDDQSITRTVACGCA--RCKGPLPDDCHBOCAAGCTPRKSDCLACH 257
 DB 165 EYG-CWEGAGHNCQRESKLNCSPQSGCGHCFGKPRECHLFCAGGCTGTPQSDCLACN 223
 QY 258 FHNSGICELHCPALVTYNDDTFESMPNPEGRYTFGASCTACPYNVLSTDVSGCTLVCP 317
 DB 224 FYDDGVCKQCECPMQYNPTNFWENPFGKAYGATVCRKP-EHLLDNAGACVAKCKX 282
 QY 318 HNOVTAEDGTORCEKCSKPCARVCYGLGMOYITANSKRTGTTELEFAGCKKIFGSLAF 377
 DB 283 GKMPQNSE-----CVPCKGVCPKTCPEGBI-----VHSDWIG-----NYKCTIIESLEIL 329
 QY 378 PESFDDPASNT-----APLOPOLOVFELEBITGYLYISAMPDLPDLSYFQNLQ 429
 DB 330 DQSFQGFQVYNNFSGPRYIKIDPRLEVFSTVKEITGFINQAHHPFTLLNRYFRLE 389

```

QY 430 VIRGILLINNGAY-SLTQLGGLISWLGRLSELGSLALIHNNHLCGFVTVPMWDLFRN 488
DB 390 VVGRQLKENIFASVTVYKTSLSKLELKLKRVNNGSIVLENSDLCTFVEDIDMSLKKS 449
QY 489 PHQALHTANRPEDECEVGEGLACHOLCARGHCMWPGPTQCVCNCSQPLRGQCEVECEVLQ 548
DB 450 SDHEVMVQKNNATECHEGEGECGSCSKAGCMWKGPEQCLCNVAVYKXKCLDSCX--- 506
QY 549 GLPREY-VNANHCLPCHPECCOPONGSVTCFEPADQCAACHYDHPFCVACRCP----- 601
DB 507 SLFRLYSVDSKTCDDCHQOECD-----FCYGNEDNCGSCNNVXDRFCAACEPTTHAM 561
QY 602 ---FNNFVFFMLRVPK-----VSASHLE-----PD----- 624
DB 562 NGTCINCHKTCCVGRPRDTIAPDGCISCDKAIIGSDAKIERCLMKDESCDDGYSDYVL 621
QY 625 -EEG----- 627
DB 622 QEEGFLKQISGKAVCRKCHPRCKCTGYGFHEGFCOECTGYKGECEDECPDPFVANE 681
QY 628 -ACOPCPINCT-----HSCVDL-----DP-----KGGPAEQ----- 652
DB 682 TRICLPCHECRGCHGLADHHECRNLKLFEGDEYDATTFTCVSNCPASHPRYRFOEA 741
QY 653 -----RASPLTISVASAVGILLVVLGVVFGI---LIKROQKIRRYTM 693
DB 742 GKIGPYCSADSMOSGLRIEPOTVKIVWGSVMALLICVFGIAGFVLFSSHKNKDXAVKM 801
QY 694 RLLQETELVPLRPSGAMPNOAMRIKETELRKVYLGAGAGTAYKGIWPDGKNV 753
DB 802 TMLAGCEDESEPLRPSNVGPNLTKRILKEAIRGGVLMGAGARFRKGVMMEGESVK 861
QY 754 IPAIVKLVRENTSPKANKEILDEAYVWAGVSPVSRLLGICLTSTVOLVTLMPYGLL 813
DB 862 IPAIVKLVRENTSPKANKEILDEAYVWAGVSPVSRLLGICLTSTVOLVTLMPYGLL 921
QY 814 DHVRENRLSGLSODLWCMQIAGKMSYLEDVRLVHDLARNLVYSPNHVKTIDEGLA 873
DB 922 DYVRNKKDKISSKALLNWSQIARGMALYERLVRHDLARNLVYTPSCVKTIVGGLA 981
QY 874 RLIDIDEYHADGGKVPKMMALLESILRRFTQSDVWSYGVTVWELMELFGAKPYGIP 933
DB 982 KLIDFDSDEYRAGSKMPTKMLALECIRHVFSTKSDVMARGLITWELTIGARPYENVP 1041
QY 934 AREIPDLKEBERLPORPITCTIDYVMTWKCMYDSECRPFRELVSEFSRMAPDPORFV 993
DB 1042 AKDYBELIEIHKLPQDPICSLDYVCTILSCWVLDADARPTFKOLAEFKAERKADPERYL 1101
QY 994 VIONEDLGPASPLDSTFYRSLLEDDEMDLV----- 1024
DB 1102 MI-----FGDKFMRLPSTYNODEKDLRTLAAYMAAALAAAGASNVVPESTIA 1152
QY 1025 DAEYLVPOGFCFCDPAFAGAGVNHHRSSSTRSGGDLTLGLPSESEEPARS----- 1079
DB 1153 ETDEYLCPTKTRPSIMLPQPSA-----VEBS-DEMPKSLRYCK 1188
QY 1080 -PLAP---SEGAGSVDFDGLGWSAKGLQSLPTHDSPLGRYSEDPVPLPSTDDYVA 1135
DB 1189 DPLKPDDETDGHEV---GVGSR-----LNLPLDEDDITLM 1222
QY 1136 PLTCSPPQPEYVQDPVRPQPSFREGPLPAARPAATLERAKTSLPGKNGVYCVAFGG 1195
DB 1223 P-TCOSQ---NOS-----TPG-----YMDLIGVPA 1243
QY 1196 AVENPEYL-----TPGGAAPQHPHPPAPSPAPFNDNLVYMODPPEKARAPSTFKGT 1246
DB 1244 SVDNPEYLMGSTQALAGAGSGM--PHTPP-----PNTTNGM 1280
QY 1247 PTAENPE 1253
DB 1281 PTHQHSQ 1287

```

```

RESULT 8
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10465918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor."
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR SMART; SM00261; Fur_1_domain; 1.
SQ SEQUENCE 419 AA; 45472 MW; FBC1BE347E2D030C CR664;

Query Match 27.3%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 3,3e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 METALCRGGLLALLPRGAATQVCTGDMKLRPASFTLDMRLHYOGCCVYQVNL 60
DB 1 METALCRGGLLALLPRGAATQVCTGDMKLRPASFTLDMRLHYOGCCVYQVNL 60
QY 61 ELTYLPTNLSLFLDIOGVQGVLLAHNQVQVPRQRLIRYRGQLFEDNALAVLNG 120
DB 61 ELTYLPTNLSLFLDIOGVQGVLLAHNQVQVPRQRLIRYRGQLFEDNALAVLNG 120
QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNQLA 180
DB 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNQLA 180
QY 181 LTLIDTPNRBACHPGSPMKSGRCWGESSESDQSILRTYACAGGACRCKGPLPTDCHEGC 240
DB 181 LTLIDTPNRBACHPGSPMKSGRCWGESSESDQSILRTYACAGGACRCKGPLPTDCHEGC 240
QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTRESMPNREGRTYFGASCYTACP 300
DB 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTRESMPNREGRTYFGASCYTACP 300
QY 301 YNLTSTDVSGCTLVGCIHNOEYTAEGTGRCKSGSPCARVYCYGL 345
DB 301 YNLTSTDVSGCTLVGCIHNOEYTAEGTGRCKSGSPCARVYCYGL 345

RESULT 9
ID Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```


RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC027080; AA027080.1; -
 KW Hypothetical protein.
 SO SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 2,6e-123;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 895 MALESLRRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 954
 DB 1 MALESLRRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60
 QY 955 IDVYIMVWCWIMIDSECRPRFRELVSFSSRMARDQRFVVIQNEDLGPASPLDSTFYRSL 1014
 DB 61 IDVYIMVWCWIMIDSECRPRFRELVSFSSRMARDQRFVVIQNEDLGPASPLDSTFYRSL 120
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1074
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1075 EAPRSPPLASSEGAGSDVFDGDLGMAAKGLQSLPTHDSPLQRYSEDPVLPSETDGY 1134
 DB 181 EAPRSPPLASSEGAGSDVFDGDLGMAAKGLQSLPTHDSPLQRYSEDPVLPSETDGY 240
 QY 1135 APLTCSPOPEYVNOVDVPRQPPSPREGPLPARPAGATLTERAKTILSPGKGVKDVFAFG 1194
 DB 241 APLTCSPOPEYVNOVDVPRQPPSPREGPLPARPAGATLTERAKTILSPGKGVKDVFAFG 300
 QY 1195 GAVENPEYLTPOGGAAPDPFPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEY 1254
 DB 301 GAVENPEYLTPOGGAAPDPFPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEY 360
 QY 1255 LGLDVVPY 1261
 DB 361 LGLDVVPY 367

RESULT 10
 Q8WVVO PRELIMINARY; PRT; 412 AA.
 AC Q8WVVO;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 44.7 KDa protein.
 GN PP3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF18349; AA55856.1; -
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc_K; 1.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397FF327D2BC CRC64;
 Query Match 24.8%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 4,4e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 895 MALESLRRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 954
 DB 1 MALESLRRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60
 QY 955 IDVYIMVWCWIMIDSECRPRFRELVSFSSRMARDQRFVVIQNEDLGPASPLDSTFYRSL 1014
 DB 61 IDVYIMVWCWIMIDSECRPRFRELVSFSSRMARDQRFVVIQNEDLGPASPLDSTFYRSL 120
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1074
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1075 EAPRSPPLASSEGAGSDVFDGDLGMAAKGLQSLPTHDSPLQRYSEDPVLPSETDGY 1134
 DB 181 EAPRSPPLASSEGAGSDVFDGDLGMAAKGLQSLPTHDSPLQRYSEDPVLPSETDGY 240
 QY 1135 APLTCSPOPEYVNOVDVPRQPPSPREGPLPARPAGATLTERAKTILSPGKGVKDVFAFG 1194
 DB 241 APLTCSPOPEYVNOVDVPRQPPSPREGPLPARPAGATLTERAKTILSPGKGVKDVFAFG 300
 QY 1195 GAVENPEYLTPOGGAAPDPFPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEY 1254
 DB 301 GAVENPEYLTPOGGAAPDPFPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEY 360
 QY 1227 YWMD-ODPPER-----GAPSTFGPTTEN 1251
 DB 361 YWMD-ODPPER-----GAPSTFGPTTEN 410

RESULT 11
 Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPROTEIN.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 NC NCB1_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Venustrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60725.1; -
 DR HSP; P03322; 1A6S
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR004028; Retro_M.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF02813; Retro_M; 1.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; Tyrc_K; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Tyrosine-kinase.
 SO SEQUENCE 729 AA; 80649 MW; 84D2F6914EFD1D63 CRC64;

Query Match 24.5%; Score 1679; DB 15; Length 729;
 Best Local Similarity 53.8%; Pred. No. 2,6e-118;

Matches 357; Conservative 71; Mismatches 119; Indels 116; Gaps 17;

QY 569 POKSVTCFGEADQCACAHYKDPFCVACRCP-----NNFTYSFMLRVKVSASHLEPDP 624
 DB 141 PEETATPRTGP--DHCKCAHAFIDGPHCVACRCPAGVIGENDTL-VWKYA-----D 187
 QY 625 EEGACQPCPINCTHSCVDLDKCGPAEQRASPLTISAVV-GILLVVLGVVGLIKR 683
 DB 188 ANAVCOLCHPNCIRGCGKPGLEGCP---NGSKTPSIAGVVGILCLVVGILGLYLRR 244
 QY 684 RQCKIRKRYMRRLIQETELVEPLTPSCAMPNQAOMRLKETELRKVKVLSGAFVYK 743
 DB 245 R-HIVRKRTLRRLQERELVEPLTPSCAMPNQAOMRLKETELRKVKVLSGAFVYK 303
 QY 744 IMIPDGENVKIPVAIKVLRNTSPKANKELIDEAVYVAGSPVSVLLGICLTSTVQLV 803
 DB 304 LMPEGEKVKIPVAIKELRENTSPKANKELIDEAVYVAGSPVSVLLGICLTSTVQLV 363
 QY 804 TQMPYGCCLLDHVENRGRIGSODLNMCMQIAKMSYLEDVRLVHRDLAARNVLYKSPN 863
 DB 364 TQMPYGCCLLDYIREHNDNIGSYLLNMCMQIAKMSYLEDVRLVHRDLAARNVLYKTPQ 423
 QY 864 HVKTTDFGLARLLIDETEHADGKVPKIMMALESILRRRFTHOSQVMSYGVTVWELMT 923
 DB 424 HVKTTDFGLARLLIDETEHADGKVPKIMMALESILRRRFTHOSQVMSYGVTVWELMT 483
 QY 924 FGAPYDGIAPAREIPDLLEKGERLPORPCTIDVYMIWVKCMIDSECRPFRELVSEFS 983
 DB 484 FGSKPYDGIAPAREIPDLLEKGERLPORPCTIDVYMIWVKCMIDSECRPFRELVSEFS 543
 QY 984 RMAADPQRFVYIQ-NEDLGPRASPLDSTFYRSLEDDMDGLVDAEYLVPQCGFCDDPA 1042
 DB 544 RMAADPQRFVYIQ-NEDLGPRASPLDSTFYRSLEDDMDGLVDAEYLVPQCGFCDDPA 598
 QY 1043 PGAGMGNHRRSSSTRSGGDLTLGLEPSEEARPSPL-----APSEGAGDVFDGDLG 1097
 DB 599 -----NSPST-----SRTPLLSLSLATSNNAATNCID----- 625
 QY 1098 MGAAGLQSLPTHDPSLQRYSEDPTVPLPSET--DGVAVLPTCSPOPEYVNOQDVAPQP 1155
 DB 626 ---RNGCGHVRSDSEFVQRYSDPTGNFLEESIDDFL-----PAPEYVNO--LMPKK 673
 QY 1156 PSPEGGPLPARAPGATLERAKTLSPGKNGVYKDF-----AFGAVENPEYL 1200
 DB 674 PS-----TAMVQNIYNNISLTALSKLPMDSRYQNSHSTAVDNP 712
 QY 1201 EYL 1203
 DB 713 EYL 715

RESULT 12

Q66714 PRELIMINARY; PRT; 567 AA.
 AC Q66714;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791.
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL, S69372; AAC60727.1; .

DR HSSP; P11362; 1FGK.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW Tyrosine-protein kinase.
 FT
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 24.5%; Score 1677; DB 15; Length 567;
 Best Local Similarity 54.4%; Pred. No. 2.5e-118;
 Matches 356; Conservative 69; Mismatches 113; Indels 116; Gaps 17;

QY 578 GPEADQCACAHYKDPFCVACRCP-----NNFTYSFMLRVKVSASHLEPDEGACQPCP 633
 DB 1 GP--DHCKCAHAFIDGPHCVACRCPAGVIGENDTL-VWKYA-----DANAVCOLCH 47
 QY 634 INCTHSCVDLDKCGPAEQRASPLTISAVV-GILLVVLGVVGLIKRQCKIRKT 692
 DB 48 PNCIGCKGKPGLEGCP---NGSKTPSIAGVVGILCLVVGILGLYLRR-HIVRKRT 103
 QY 693 MRLLQETELVEPLTPSCAMPNQAOMRLKETELRKVKVLSGAFGVYKGIWIPDGENV 752
 DB 104 LRRLLQERLVEPLTPSCAMPNQAOMRLKETELRKVKVLSGAFGVYKGIWIPDGENV 163
 QY 753 KIPVAIKVLRNTSPKANKELIDEAVYVAGSPVSVLLGICLTSTVQLVQMPYGCCL 812
 DB 164 KIPVAIKVLRNTSPKANKELIDEAVYVAGSPVSVLLGICLTSTVQLVQMPYGCCL 223
 QY 813 LDHVENRGRIGSODLNMCMQIAKMSYLEDVRLVHRDLAARNVLYKSPNKTITPGL 872
 DB 224 LDYIRHKNINISQYLLNMCMQIAKMSYLEDVRLVHRDLAARNVLYKTPQHVKITDPGL 283
 QY 873 ARLLIDETEHADGKVPKIMMALESILRRRFTHOSQVMSYGVTVWELMTFGAPYDGI 932
 DB 284 AKLLGADKEKETHAEGKVPKIMMALESILRRRFTHOSQVMSYGVTVWELMTFGAPYDGI 343
 QY 933 PAREIPDLLEKGERLPORPCTIDVYMIWVKCMIDSECRPFRELVSEFSMAADPQRF 992
 DB 344 PASELSSVLEKGERLPORPCTIDVYMIWVKCMIDSECRPFRELVSEFSMAADPQRF 403
 QY 993 VYIQ-NEDLGPRASPLDSTFYRSLEDDMDGLVDAEYLVPQCGFCDDPAAGAMVH 1051
 DB 404 LVIQGERKHLSPDTSKRYRTLMEEDMEDIVDAEYLVPQCGF----- 449
 QY 1052 RHRSSSTRSGGDLTLGLEPSEEARPSPL-----APSEGAGDVFDGDLGGAAGLQSL 1106
 DB 450 ---NSPST-----SRTPLLSLSLATSNNAATNCID-----RNGG 481
 QY 1107 LPTHDPSLQRYSEDPTVPLPSET--DGVAVLPTCSPOPEYVNOQDVAPQPSPREGPLP 1164
 DB 482 HVRREDSFYQRYSDPTGNFLEESIDDFL-----PAPEYVNO--LMPKKSS----- 526
 QY 1165 AARPAAGATLERAKTLSPGKNGVYKDF-----AFGAVENPEYL 1203
 DB 527 -----TAMVQNIYNNISLTALSKLPMDSRYQNSHSTAVDNP 566

RESULT 13

Q64895 PRELIMINARY; PRT; 962 AA.
 AC Q64895;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Gag v-erbA, v-erbB protein.
 GN Gag v-erbA, v-erbB.
 OS Avian erythroblastosis virus.

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 NCBI_TaxID=11861;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90206603; PubMed=1969616;
 RA Brinkman A., Jackson J., Bishop J.M., McCarty D.J., Schlitzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 transforming potential of the oncogene v-erb-B-";
 RL Oncogene 5:15-24(1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL: X52209; CA36459.1; -
 DR EMBL: X52211; CA36459.1; JOINED.
 DR HSSP: P10828; NTL.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000536; Hormone_rec_1ig.
 DR InterPro: IPR001723; Stdhm_recptor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR001628; Znf_Cstereoid.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00105; Zf-C4; 1.
 DR PRINTS: PR00398; STRDHOMONER.
 DR PRINTS: PR00047; STROIDIFINGER.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR ProDom: PD000035; Znf_Cstereoid; 1.
 DR SMART: SM00430; HOL1; 1.
 DR SMART: SM00239; TykC; 1.
 DR SMART: SM00399; Znf_C4; 1.
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
 Transcription regulation; Transferase; Tyrosine-protein kinase;
 Zinc-finger.
 KW SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;
 SQ

Query Match 23.6%; Score 1612.5; DB 15; Length 962;
 Best Local Similarity 50.9%; Pred. No. 4.2e-113;
 Matches 357; Conservative 69; Mismatches 143; Indels 133; Gaps 20;

QY 541 VEECRVLQGLPRE-YVNAR-HCLP-----CHEECQ 568
 DB 154 IEKQESYLLAFEHYINVRKHNIHFWKSLMKLVADLMIGAYHASFILMKYCEPTELS 413
 QY 569 PONGSVTCFGEADQCVACAHYKDPFCVACPF-----NFTVSFWLRVKVVSASHLEPD 624
 DB 414 PQE-----VGP--DHCKCAHFIDGPHCVACAPGAVIGENDTL-VWKYA-----D 455

QY 625 EEGACQCPINCTHSCVDLDDGCPAEQASPLTSIVSAVY-GILLVVLGVVFGILIKR 683
 DB 456 ANAVCQCHNCNCTGCGGPGLEGCP---NGSKTPSIAGVVGGLICLVVGLGILYLR 512

QY 684 ROKIKRYTWRLLOETELVEPLTPSGAMPNOAOMRIKTELKRYKVLGSGAFVYK 743
 DB 513 R-HIVKRTLRRLQERELVEPLTPSGEAPNOAHLRIKTEFEKRVVLGFGAFVYK 571

QY 744 IWPIDGENVAKIPVAIKVIRENTSPKANKELIDEAYVAVGSPYVSLGICLSTVQVL 803
 DB 572 IWPIDGENVAKIPVAIKVIRENTSPKANKELIDEAYVAVGSPYVSLGICLSTVQVL 631

QY 804 TQLMFYGCLLDHVENRGSLGSDLLMCMQIAKMSYLEDVRLVHDLAARVAVLSP 863
 DB 632 TQLMFYGCLLDHVENRGSLGSDLLMCMQIAKMSYLEDVRLVHDLAARVAVLSP 691

QY 864 HVKILDFGLARLIDIDEYHADGKVPKIMMALESTLRRTFHOSDWSYGVTVWELNT 923
 DB 692 HVKILDFGLARLIDIDEYHADGKVPKIMMALESTLRRTFHOSDWSYGVTVWELNT 751

QY 924 FGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYMIWYKMMIDSECRPRFELSEFS 983
 DB 752 FGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYMIWYKMMIDSECRPRFELSEFS 811

QY 984 RMAADPQRFVVIC-NEDLGASPLDSTFYRSLEDDDMGLVDAEEYLVPOGFCFDPDA 1042
 DB 812 KVARDDPRLYLVQGDERRMLPSPDTSKFRITLMEEDMEDIVADAEYLVPHQGF----- 866

QY 1043 PGAGVNHRRHSSSTRSGGDTLLGLEPSEEBAPSPPLASBAGSDVFDGDLGMAK 1102
 DB 867 -----NSPST-----SKTPLLSSISATSN-----NSATK 890

QY 1103 GLQSLPTHPDPSLPORVSEDPVPLPSETGVYVAPLTCSPQEFYVNOPDVRRPOPSPREG 1162
 DB 891 CIDRNGH-----PREDGFL-----PAPEYVNO-LMPKPFSTAMVQ 926

QY 1163 LPARAPAGAT-IERAKTLSPGKNGVYKDVFAFGAVENPEYL 1203
 DB 927 NQIYNYISLTALSKLEPMSDRYQN-----SHSTAVDNPEYL 961

RESULT 14
 Q85468 ID Q85468 PRELIMINARY; PRT; 545 AA.
 AC Q85468;
 DT 01-NOV-1996 (TREMELREL. 01, Created)
 DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
 DT 01-MAR-2002 (TREMELREL. 20, Last annotation update)
 DE Avian erythroblastosis virus (Ts34) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 NCBI_TaxID=11861;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=88217326; PubMed=2897102;
 RA "Scotting P., Vennistrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation.";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL: X06943; CA30024.1; -
 DR HSSP: P13562; ITRK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; TykC; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR ATP-binding; Transferase; Tyrosine-protein kinase.
 KW SEQUENCE 545 AA; 60899 MW; 140DC8ECCA0F8AF4 CRC64;
 SQ

Query Match 23.4%; Score 1604; DB 15; Length 545;
 Best Local Similarity 53.9%; Pred. No. 8e-113;
 Matches 344; Conservative 66; Mismatches 122; Indels 106; Gaps 17;

QY 578 GREADQCVACAHYKDPFCVACPF-----NFTVSFWLRVKVVSASHLEPDGACQPCP 633
 DB 1 GP--DHCKCAHFIDGPHCVACAPGAVIGENDTL-VWKYA-----DANACQLCH 47

QY 634 INCTHSCVDLDDGCPAEQASPLTSIVSAVY-GILLVVLGVVFGILIKRPOQIKRYT 692
 DB 48 PNCRTGCKGPGLEGCP---NGSKTPSIAGVVGGLICLVVGLGILYLR-R-HIVKRT 103

QY 693 MRLLQETELVEPLTPSGAMPNOAOMRIKTELKRYKVLGSGAFVYKGIWIDGNNV 752
 DB 104 LRLLQERELVEPLTPSGEAPNOAHLRIKTEFEKRVVLGFGAFVYKGLMPEGERV 163

QY 753 KIPVAIKVIRENTSPKANKELIDEAYVAVGSPYVSLGICLSTVQVLQTLMPYGL 812
 DB 164 TTPVAIKVIRENTSPKANKELIDEAYVAVGSPYVSLGICLSTVQVLQTLMPYGL 223

QY 813 LDHVENRGSLGSDLLMCMQIAKMSYLEDVRLVHDLAARVAVLSPNHVKTDTGCL 872
 DB 224 LDYIREHKNIGSQYLLMCMQIAKMSYLEDVRLVHDLAARVAVLSPDVKITDGL 283

QY 873 ARLLDDETEYHADGKVPKIMMALESILRRRFTHQSDVMSGYVTWELMTFGAPYDGI 932
 DB 284 AKQGLADEKYEYABEGKVPKIMMALESILHRYTHQSDVMSGYVTWELMTFGSKPIYGI 343
 QY 933 PARETIDLEKGERLPDPICITIDYMTWKCMTIDSCRRPRELVESEFRMADPQRF 992
 DB 344 PASEISVLEKGERLPDPICITIDYMTWKCMTIDSCRRPRELVESEFRMADPQRF 403
 QY 993 VVIQ-NEDGPASPLDSTFFRSLBEDDMGDLVDEEYLVVQOQFCFCDPAPGAGMWH 1051
 DB 404 LVIOQDERMHLPSPTDSKFRYRMTMEEDMEDIVDADELIVHQQF----- 449
 QY 1052 RHRSSTNSGGDLTLGLEPSEEAARSP-----APSEGASGVFDGDLGMAKGLQS 1106
 DB 450 -NSPST-----SRTPLLSLSATSNMNSATNCDRNG----- 480
 QY 1107 LPTHDPSPLOQRSEDPYPLPSPETGVYAPLTCSPQRYNQPDPVPPSPREGPLPA 1166
 DB 481 -H-----PVREDGFL-----PAPRYVQO-LMPKKSTANVQIQIY 513
 QY 1167 RPAGAT-LEBAKTLSPGNKGVKDYAFAGAVENPEYL 1203
 DB 514 NYISLTALSKLPMDSRYQN-----SHSTAVDNEYL 544
 RESULT 15
 Q9WVF5 PRELIMINARY; PRT; 655 AA.
 AC Q9WVF5; 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
 DE
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB Taxid:10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C., Lampfand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Mable N.J.;
 RA "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J., Sinselar C.S., Pearsall R.S., Green P.J., Yee D., Lampfand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Mable N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms.";
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RA MEDLINE=21085660; PubMed=11217851;
 RA Kawata T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F., RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L., RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AF124513; AAD4149.1; -
 DR EMBL: AF275366; AAG28047.1; JOINED.
 DR EMBL: AF275366; AAG28047.1; JOINED.
 DR EMBL: AF275365; AAG28047.1; JOINED.
 DR EMBL: AK004944; BAB23688.1; -
 DR EMBL: AK004883; BAB23641.1; -
 DR EMBL: AK004911; BAB23662.1; -
 DR MGI: MGI:95294; Egrf.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; Fv; 3.
 KM Receptor.
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;
 Query Match 21.5%; Score 1468.5; DB 11; Length 655;
 Best Local Similarity 43.9%; Pred. No. 1,9e-102;
 Matches 286; Conservativity 90; Mismatches 234; Indels 41; Gaps 11;
 QY 11 LIALALPFGAA--STOVCTTDMKRLPASPEHLMRLHYGCGVGNLEFLYPTN 68
 DB 14 LITLALCAGALBEKRYCQGSNRLTQLGFEDHFLSLDRMNNCEVYGNLEITYVOR 73
 QY 69 ASLSPFDIDIOGVYLIHNOYRQVPLRLVRGTQFEDYALAVDNGPLNNTTP 128
 DB 74 YDSFLTKTIOEVAGVLIHNTYRPLENTQIRGNALYENTYALALISN----- 124
 QY 129 VTGASPGRLREIQLSLTEILKGVLIORNPOLCYODTILMKDI---FKKNQLALTL 184
 DB 125 -YGTNRTGLRELPMRNLOETLIGAVFNSNPILCNMDTIQMRDIYGVNFSNMSMDL--- 180
 QY 185 DTRNSAACHSCSMKCGSRWSSSESDQSLRTVAGGCA-RCKRPLPTDCHDCCAG 243
 DB 181 -QSPSSCPKCDSCNPGSCWGGEBNCCQLTKITICAOQSHCRGRSPDCCNQCAG 239
 QY 244 CTGPKHSDCIACLFHNSGICELHCPALVTYNTDPESMPNPEGRYTFGASCYTAAPY 303
 DB 240 CTGPRSDCLVQCKFDDEATCKDTCRPLMLYNPTTYQMDVNPGRKVSFGATCYKCPRY 299
 QY 304 LSTDVSGCTVLCPLHNOEYTAEDGTORCEKSPCARVCGYGLMOYIKANSKFTGLE 363
 DB 300 VVTDHGSQVACCPDYEV-EEDGIRKCKCDPRKVCNNGIGIEFK-DTSLINATNIK 357
 QY 364 -FAGCKKIFGSLAFLESPDPDPASNTAPLOPQOVFETLEITVLYISAMPDPLD 422
 DB 358 HFXYCAISDHLILVAKRGDSFTTPPDLRELILTKVETITFFLLIQAMPDWTDL 417
 QY 423 SVFQNTQVTRGRIHNGAVSLTQLGIGISMLGRLSRLDGSGIALIHNTHLCEYTV 482
 DB 418 HAFENLEIRGRTKHGQFSLVAVGINTISLGRSLKEISDGDVILISGRNLCYANTIM 477
 QY 483 DQLEFRPHOALHTARPPDECEVGEGLAHO-CARHCGMGPGTQCVNCSQFLRQECVE 542
 DB 478 KKLFGFNPQTKTKMNNRAEKDCAVNVHVCNPLCSSGCGPBRDVCSCQVSRGECVE 537
 QY 543 ECRVQLGPREVYNAHCLCPHECCQPNQSVTCFQPREADQVACAHYDPFCVAPCF 602
 DB 538 KCVILGEBPREFVNESECICPHECLPQMMNITCTGRPDNCIQCAHYIDGPHCVKTC 597
 QY 603 -----NNFTVSPFLRKPXKASHLPEDEGAQCPGICINTHSCVDLDIDGCG 648
 DB 598 GIMGENNTLV--WKYA-----DANNVCHLCNANTCYAGGGLGGC 636

Tue Jul 22 12:41:42 2003

seq4-579-593-14.rpt

Page 13

Search completed: July 22, 2003, 09:01:39
Job time : 54.5887 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds

(without alignments)
4403.399 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 1 MELALCRWGLLALLRPGA.....TFKCTPTAENPEYIGLDVIV 1261

Sequence: BL0SUM62

Scoring table: Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

```
Database :
1: /Geneseq_101002:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6627	96.9	1255	21	AA92620	Human heregulin 2
2	6627	96.9	1255	22	AAE12130	Human tyrosine kin
3	6627	96.9	1255	22	AAE60167	HER2 transgene pla
4	6627	96.9	1255	23	AAU74545	Human HER2 (erbB2)
5	6621	96.8	1255	17	AAU01111	HER-2/neu protein.
6	6621	96.8	1255	20	AA92406	Human HER-2/neu on
7	6621	96.8	1255	21	AAE21198	Human HER-2/neu pr
8	6621	96.8	1255	21	AAE84780	Amino acid sequenc
9	6621	96.8	1255	22	AAE85458	Human HER-2/neu pr
10	6621	96.8	1255	22	AAE88267	HER2/neu amino aci

11	6621	96.8	1255	23	AAE24067	Human Her-2 protei
12	6621	96.8	1255	23	AAE20479	Human Her-2/neu pr
13	6621	96.8	1255	23	AAE51143	Human Her-2/neu on
14	6621	96.8	1255	23	AAU77114	Human Her-2/neu po
15	6621	96.8	1255	14	AAE39568	Sequence of c-erbB
16	6457	94.4	1223	23	AAU98923	Human breast cance
17	6304	92.2	1200	21	AAE21208	Human HER-2/neu pr
18	5837.5	85.3	1256	21	AAE21199	Rat HER-2/neu prot
19	5837.5	85.3	1256	23	AAE51144	Rat HER-2/neu onco
20	5810.5	84.9	1256	21	AAE21206	Mouse Her-2/neu pr
21	5810.5	84.9	1256	22	AAE62860	Amino acid sequenc
22	5810.5	84.9	1256	23	AAE51151	Mouse Her-2/neu on
23	4707	68.8	919	21	AAE21203	Human HER-2/neu fu
24	4707	68.8	919	23	AAE51148	Human HER-2/neu fu
25	3660.5	57.9	920	23	AAE51152	Mouse Her-2/neu ex
26	3660.5	57.9	926	23	AAE51153	Mouse Her-2/neu ex
27	3591	52.5	712	21	AAE21204	Human HER-2/neu fu
28	3591	52.5	712	23	AAE51149	Her-2/neu extracel
29	3445	50.4	782	18	AAE19764	Her-2-GF-CGF immuno
30	3443	50.3	653	21	AAE21200	Extracellular HER-
31	3443	50.3	653	23	AAE51145	Human Her-2/neu on
32	3405	49.8	645	22	AAE60408	Human ErbB2 oncopr
33	3405	49.8	645	22	AAE61593	Human ErbB2 extrac
34	3340	48.8	951	21	AAE44993	DC9scFv-erbB2EC fu
35	3237	47.3	624	11	AAE08822	Extracellular port
36	3110	45.5	1210	21	AAE19255	Amino acid sequenc
37	3110	45.5	1210	21	AAE50616	Human EGF receptor
38	3110	45.5	1210	23	AAE23019	Human Her-1 protei
39	3110	45.5	1210	23	AAE50768	Human epidermal gr
40	3108	45.4	1210	22	AAE68420	Amino acid sequenc
41	3084	45.1	583	23	AAE20483	Human protein for
42	3084	45.1	587	23	AAE20481	Human protein for
43	3083	45.1	589	23	AAE20482	Human protein for
44	3083	45.1	600	23	AAE20484	Human protein for
45	3069	44.9	1210	23	AAE51768	Human epidermal gr

ALIGNMENTS

RESULT 1	AA92620	standard; Protein; 1255 AA.
ID	AA92620	
XX	AA92620;	
AC		
XX		
XX	10-AUG-2000 (first entry)	
XX		
DE	Human heregulin 2 (Her2).	
XX		
KW	Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
XX		
OS	Homo sapiens.	
XX		
XX		
PH	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT	Region	/label= Cysteine_rich_domain
FT	Region	210..224
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	250..264
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	324..483
FT	Region	/label= ligand_binding_domain
FT	Region	325..339
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	369..383
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	465..479
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	484..623
FT	Region	/label= Cysteine_rich_domain
FT	Region	579..593
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	624..654
FT	Region	/label= Transmembrane_domain
FT	Region	632..652
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	653..667
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	655..1010
FT	Region	/label= Tyrosine_kinase_domain
FT	Region	661..675
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	695..709
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	710..730
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	1011..1235
FT	Region	/label= C-terminal_domain
FT	Region	
PN	WO200020027-A2.	
PD	13-APR-2000.	
XX		
PF	05-OCT-1999;	99WO-DK00525.
XX		
PR	05-OCT-1998;	98DX-0001261.
PR	20-OCT-1998;	98US-0105011.
XX		
XX	(MEBI-) M & E BIOTECH AS.	
XX		
PI	Steina L, Mouritsen S, Nielsen KG, Haaning U, Leach D, Dalum I,	
PI	Gautam A, Birk P, Karlsson G;	
XX		
DR	WPI; 2000-349917/30.	
DR	N-PSDB; AAA09455.	
XX		
PT	Inducing immune responses to weakly immunogenic, tumor associated	
PT	peptide antigens for the treatment of breast and prostate cancer	
PS	Claim 62; Page 193-198; 220pp; English.	
XX		
XX	This is the human heregulin 2 (Her2) sequence. Immunogenic analogues	
CC	Her2 can be used in the claimed method as an adjuvant to induce a	
CC	response. Subdominant CTL epitopes, antibody binding regions and	
CC	cysteine residues involved in disulfide bonds are preserved in the	
CC	immunogenized forms. Regions suitable for the insertion of foreign	
CC	helper epitopes were identified (see features table). The method	

Query	Match	96.9%	Score 6627	DB 21	Length 1255
Best Local Similarity	97.0%	Pred. No. 0			
Matches 1227	Conservative	6	Mismatches 18	Indels 14	Gaps 2
CC	is used for inducing immune responses against weakly immunogenic				
CC	cell-associated peptide antigens (PA) such as those associated with				
CC	cancers (self-proteins), e.g. human prostate specific membrane antigen				
CC	(PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).				
CC	The method comprises effecting simultaneous presentation by antigen				
CC	producing cells (APCs) of the animals immune system of: (1) at least 1				
CC	CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1				
CC	B-cell group derived from the cell-associated PA; and (2) at least 1				
CC	first T helper cell group which is foreign to the animal. Analogues of				
CC	human PSM, human Her2 and human/murine FGF8b comprising a substantial				
CC	part of all known and predicted CTL and B-cell epitopes of the respective				
CC	PA and including at least one foreign T helper epitope are also claimed.				
CC	The method is used to treat prostate, prostate/breast or breast cancer				
CC	when the PA is human PSM, FGF8b and Her2, respectively.				
xx					
xx	Sequence 1255 AA:				
QY	1 MELAALCMWGLILLPFGASTQVCTGTDKRLPLPASPEHMLRLHYGGCCVQVQGNL 60				
DB	1 MELAALCMWGLILLPFGASTQVCTGTDKRLPLPASPEHMLRLHYGGCCVQVQGNL 60				
QY	61 ELTYLPTNASLSPFODIQEVQGYVLIAHQVQVPLQRLIRVGTQJLEDVYALAVLDNG 120				
DB	61 ELTYLPTNASLSPFODIQEVQGYVLIAHQVQVPLQRLIRVGTQJLEDVYALAVLDNG 120				
QY	121 DPLNNTPTVTCASFGGLRELOLRSLTELKGGVLIQRNPQLCYDTILMKDIFHNKQOLA 180				
DB	121 DPLNNTPTVTCASFGGLRELOLRSLTELKGGVLIQRNPQLCYDTILMKDIFHNKQOLA 180				
QY	181 LTLIDINRSRACHPCSPCKSKRCMBGSESSCCSITRIVCAGGCAKCGPLPTDCCHQC 240				
DB	181 LTLIDINRSRACHPCSPCKSKRCMBGSESSCCSITRIVCAGGCAKCGPLPTDCCHQC 240				
QY	241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300				
DB	241 AAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300				
QY	301 YNYLSTDVSGSTIYCPILHNOEVTADGQREKSKSPARVCYGIAMGYIKANSKFIGIT 360				
DB	301 YNYLSTDVSGSTIYCPILHNOEVTADGQREKSKSPARVCYGIAMGYIKANSKFIGIT 360				
QY	361 ELLEFAGCKIKIGSLAFIPESFDGPASNTAPLQEPOLQVFFLEITIGYLIYSAMPDLP 420				
DB	361 IQEFAGCKIKIGSLAFIPESFDGPASNTAPLQEPOLQVFFLEITIGYLIYSAMPDLP 420				
QY	421 DLSVFPQNTQVIRGRILHNGAVSLTQIGISWLGIRSLRELGSGLALIHNTHTLCFHTV 480				
DB	421 DLSVFPQNTQVIRGRILHNGAVSLTQIGISWLGIRSLRELGSGLALIHNTHTLCFHTV 480				
QY	481 FMDQLFRNPQALHTANRPEDECGEGELACHQICARHCHGPGPTGCVCNSQPLRGECC 540				
DB	481 FMDQLFRNPQALHTANRPEDECGEGELACHQICARHCHGPGPTGCVCNSQPLRGECC 540				
QY	541 VEECRVTLQGLREYVNAHCLPCHPECCQPOGSGVTCFGEADQCVAAAHYKDPFCVARC 600				
DB	541 VEECRVTLQGLREYVNAHCLPCHPECCQPOGSGVTCFGEADQCVAAAHYKDPFCVARC 600				
QY	601 PFNNFTVSFMLRVKVASASHLE----PDEBACQCPRPINCHSCVDLDDKCKPAQRASP 656				
DB	601 PSFG-----VKPDLSTYMPIKPFDEBACQCPRPINCHSCVDLDDKCKPAQRASP 650				
QY	657 LTSIVSAVGLLVVLGVVFGGLIIRKQOKIRKTYMRLLQETELVEPLTPSGAMPQA 716				
DB	651 LTSIVSAVGLLVVLGVVFGGLIIRKQOKIRKTYMRLLQETELVEPLTPSGAMPQA 710				
QY	717 QMRLLKETELRKVKVLSGAFGYVKGIWIPDGENVKI PVAIKVLENTSPKANKEIIDE 776				
DB	711 QMRLLKETELRKVKVLSGAFGYVKGIWIPDGENVKI PVAIKVLENTSPKANKEIIDE 770				

QY 777 AYVAGVGSPPYVRLGLCTSTVQVLTQMPYGCCLLDHVENRGRGLSGDILINMCQIA 836
 Db 771 AYVAGVGSPPYVRLGLCTSTVQVLTQMPYGCCLLDHVENRGRGLSGDILINMCQIA 830
 QY 837 KGMSTLDEVALVHRDLAARVLYKSPNHYKITDGLARLIDIDETETHADGGVPIKMA 896
 Db 831 KGMSTLDEVALVHRDLAARVLYKSPNHYKITDGLARLIDIDETETHADGGVPIKMA 890
 QY 897 LESILRRRPTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 956
 Db 891 LESILRRRPTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 950
 QY 957 VYVIMVCKMIDSECCRRFRELVSFSRMARDQRFVIVIONEDLGASPLDSTFYRSLIE 1016
 Db 951 VYVIMVCKMIDSECCRRFRELVSFSRMARDQRFVIVIONEDLGASPLDSTFYRSLIE 1010
 QY 1017 DDMGDLVDAEELVLPQGFPCDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEA 1076
 Db 1011 DDMGDLVDAEELVLPQGFPCDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEA 1070
 QY 1077 PRSPPLAPSEAGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYAP 1136
 Db 1071 PRSPPLAPSEAGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYAP 1130
 QY 1137 LTGSPQPEYVNOVDVPPSPREGPLPAARPAAGATLERAATLSPGKNGVYVDFAFGA 1196
 Db 1131 LTGSPQPEYVNOVDVPPSPREGPLPAARPAAGATLERAATLSPGKNGVYVDFAFGA 1190
 QY 1197 VENPEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENEYIG 1256
 Db 1191 VENPEYLTPOGGAAPQHPHPAPSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENEYIG 1250
 QY 1257 LDVPV 1261
 Db 1251 LDVPV 1255

RESULT 2
 ID AAE12130 standard; Protein; 1255 AA.
 AC AAE12130;
 DT 18-DEC-2001 (first entry)
 DE Human tyrosine kinase-type receptor, HER-2.
 KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note="Antigenic epitope"
 PN WO200168677-A2.
 PD 20-SEP-2001.
 PF 16-MAR-2001; 2001MO-US40328.
 PR 16-MAR-2000; 2000US-0527487.
 PA (GEN2) GENZYME CORP.
 PI Nicolette CA;
 DR WPI; 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX

PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy; has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties
 PT
 PS Claim 4; Page 63-67; 69pp; English.
 XX

The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterized by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).

Sequence 1255 AA;

Query Match 96.9%; Score 6627; DB 22; Length 1255;
 Best Local Similarity 97.0%; Pred. No. 0;
 Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;

QY 1 MEIALCRWGLLALIPGASTVCGTDMKSLSPSPHMLRLYLGGCCVQVQNTL 60
 Db 1 MEIALCRWGLLALIPGASTVCGTDMKSLSPSPHMLRLYLGGCCVQVQNTL 60
 QY 61 ELYLPTNASLSFLQDIOEVQVYLIANQVQVPLQRLIVRGTOLEFEDYALAVLNG 120
 Db 61 ELYLPTNASLSFLQDIOEVQVYLIANQVQVPLQRLIVRGTOLEFEDYALAVLNG 120
 QY 121 DPLNNTTPTVGTASRGGLRELQSLTELKGVLIQNNPQLCYODTILMKIIFHKNOLA 180
 Db 121 DPLNNTTPTVGTASRGGLRELQSLTELKGVLIQNNPQLCYODTILMKIIFHKNOLA 180
 QY 181 LTLDITNSRACHPCSPCKGRCMGESSEDCGLTITVCGAGCARCGPLPTCCHEQC 240
 Db 181 LTLDITNSRACHPCSPCKGRCMGESSEDCGLTITVCGAGCARCGPLPTCCHEQC 240
 QY 241 AAGCTGPRGSDCLACLHFNHSGICELHCPALVTYNTDFFESMPNREGRYTFGASCVTACP 300
 Db 241 AAGCTGPRGSDCLACLHFNHSGICELHCPALVTYNTDFFESMPNREGRYTFGASCVTACP 300
 QY 301 YNYLSTVGSCTVLCPLHNOEVTAEDGTQRECKSKPCARCYGLGMEHREVAVTSAN 360
 Db 301 YNYLSTVGSCTVLCPLHNOEVTAEDGTQRECKSKPCARCYGLGMEHREVAVTSAN 360
 QY 361 ELERAGCKKIFGSLAFIPESFSDGPASNTAPLQPEQLQVFTLEITGYLYISAMPDLP 420
 Db 361 IQEFAGCKKIFGSLAFIPESFSDGPASNTAPLQPEQLQVFTLEITGYLYISAMPDLP 420
 QY 421 DLSVFNQVIRGRIILNKAVALTLQGLGISWLGRLRLREIGSLALIHNTHLCPVHTV 480
 Db 421 DLSVFNQVIRGRIILNKAVALTLQGLGISWLGRLRLREIGSLALIHNTHLCPVHTV 480
 QY 481 PMDQLFNPNHQAALHTANRPEDCVGEGALCHOLCARGHCMGPPTCCVNSQSLRQEC 540
 Db 481 PMDQLFNPNHQAALHTANRPEDCVGEGALCHOLCARGHCMGPPTCCVNSQSLRQEC 540
 QY 541 VEECRVILQGLPREYVNAARHCLPCHPEQOPNGSVTCGPEADQCVACAHYKDPFCVAC 600
 Db 541 VEECRVILQGLPREYVNAARHCLPCHPEQOPNGSVTCGPEADQCVACAHYKDPFCVAC 600
 QY 601 PFNNFTVSFMLRVFKVASLHE---PDEAGACQPCPINCSTSVLDLDDGCAEQASP 656
 Db 601 PSG-----VKPDLSTMPITWKFPEDEGACQPCPINCSTSVLDLDDGCAEQASP 650

QY 657 LTSIVSAVGLLVVVGVLGKIRKROQKIRKTYMRLLQETELVEPLTPSGAMPNQA 716
 DB 651 LTSIVSAVGLLVVVGVLGKIRKROQKIRKTYMRLLQETELVEPLTPSGAMPNQA 710
 QY 717 QMRILKETELRKVKVLSGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDE 776
 DB 711 QMRILKETELRKVKVLSGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDE 770
 QY 777 AYVWAGVSPVSSLLGCTSTVQLTQMLPFGCLLDHRENRGRGSGDLMNMCQIA 836
 DB 771 AYVWAGVSPVSSLLGCTSTVQLTQMLPFGCLLDHRENRGRGSGDLMNMCQIA 830
 QY 837 KGMVLEDEVLRLHDLAARVNLVKS PMHVKITDFGLARLLIDETEHADGKVPKIMMA 896
 DB 831 KGMVLEDEVLRLHDLAARVNLVKS PMHVKITDFGLARLLIDETEHADGKVPKIMMA 890
 QY 897 LESTLRRRPTHOSVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGEPLPQPPCTTD 956
 DB 891 LESTLRRRPTHOSVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGEPLPQPPCTTD 950
 QY 957 VYMIMVKCMMIDSECRPFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
 DB 951 VYMIMVKCMMIDSECRPFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010
 QY 1017 DDDMGDLVDAEYIVPQGGFCPPDPAPAGACGMVHRRSSSTRSGGDLTLGLEPSEBEA 1076
 DB 1011 DDDMGDLVDAEYIVPQGGFCPPDPAPAGACGMVHRRSSSTRSGGDLTLGLEPSEBEA 1070
 QY 1077 PRSPPLASEGAGSDVDFGDDLGMAKAGLSLPTHDPSPLQRYSEDDPTVPLPSETDGVYAP 1136
 DB 1071 PRSPPLASEGAGSDVDFGDDLGMAKAGLSLPTHDPSPLQRYSEDDPTVPLPSETDGVYAP 1130
 QY 1137 LTCSPQPEYVNPQDPVPPSPREGLPAARPAATLERAKTSLSPGNVGVKDVFAFGGA 1196
 DB 1131 LTCSPQPEYVNPQDPVPPSPREGLPAARPAATLERAKTSLSPGNVGVKDVFAFGGA 1190
 QY 1197 VENPEVYLTPOGGAPQHPAPAPAFDNLVYMWQDPERGAPSTFKGTPTAENPEYILG 1256
 DB 1191 VENPEVYLTPOGGAPQHPAPAPAFDNLVYMWQDPERGAPSTFKGTPTAENPEYILG 1250
 QY 1257 LDVEV 1261
 DB 1251 LDVEV 1255

RESULT 3
 ID AAB60167 standard; Protein; 1255 AA.
 AC AAB60167;
 DT 03-APR-2001 (first entry)
 DE HER2 transgene plasmid construct encoded protein.
 KM Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
 OS antibody.
 OS Homo sapiens.
 OS Synthetic.
 PN WO200100244-A2.
 PD 04-JAN-2001.
 PF 23-JUN-2000; 2000WO-US-7229.
 PR 25-JUN-1999; 99US-0141316.
 PR 16-MAR-2000; 2000US-0189844.
 PA (GETH) GENENTECH INC.
 XX Erickson S, Schwall R;

XX MPI: 2001-061962/07.
 DR N-PSDB; AAF24297.
 XX Treating tumors, particularly breast cancers, which overexpress an ErbB
 PT receptor and does not respond to an anti-ErbB antibody, comprises
 PT conjugating the antibody to a maytansinoid -
 XX
 PS Example 3; Fig 4; 92pp; English.
 CC The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
 XX
 SQ Sequence 1255 AA;
 Query Match 96.9%; Score 6627; DB 22; Length 1255;
 Best Local Similarity 97.0%; Pred. No. 0;
 Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;
 QY 1 METALCRWGLLALLPPGAASVQCTGTMKLRIPASPEHLDMLRHLVQCCVYQCNL 60
 DB 1 METALCRWGLLALLPPGAASVQCTGTMKLRIPASPEHLDMLRHLVQCCVYQCNL 60
 QY 61 ELTYLPTNASLFLDIOEVQGYVLIANNOVROVLOLRIRYGTQLFEDNYALAVLNDG 120
 DB 61 ELTYLPTNASLFLDIOEVQGYVLIANNOVROVLOLRIRYGTQLFEDNYALAVLNDG 120
 QY 121 DPLNNTPTVGSAGSRELEQRLSTELILKGVLLIQRNPOLCYQDTIMKDIFFHKNOLA 180
 DB 121 DPLNNTPTVGSAGSRELEQRLSTELILKGVLLIQRNPOLCYQDTIMKDIFFHKNOLA 180
 QY 181 LTLIDTNSRACHPCSPCKSGRCGSESEDCQSILTRVYACGACRCGPLYPTDCHEQC 240
 DB 181 LTLIDTNSRACHPCSPCKSGRCGSESEDCQSILTRVYACGACRCGPLYPTDCHEQC 240
 QY 241 AAGCTGPRKSDCLALCHFNHSGICELHCPALVYVTDFFESMPNREGYTGASCVTACP 300
 DB 241 AAGCTGPRKSDCLALCHFNHSGICELHCPALVYVTDFFESMPNREGYTGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPILHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLEBRAVTSAN 360
 DB 301 YNYLSTDVGSCTLVCPILHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLEBRAVTSAN 360
 QY 361 ELTFAGCKKIRGSLAFLPESFDGPASTAPLQBPQLQVETLEITGLVYISAMPDSL 420
 DB 361 ELTFAGCKKIRGSLAFLPESFDGPASTAPLQBPQLQVETLEITGLVYISAMPDSL 420
 QY 421 DLVFPQNTQVIRGRILHNGAVSLTLQGLIGISWLGRLSRLREISGALILHNTHLCFVATV 480
 DB 421 DLVFPQNTQVIRGRILHNGAVSLTLQGLIGISWLGRLSRLREISGALILHNTHLCFVATV 480
 QY 481 PWDQLFNPHQALLHTANRPDECEVGEGLACHQLCARHCGMPGTQCVNCSQFLRQEC 540
 DB 481 PWDQLFNPHQALLHTANRPDECEVGEGLACHQLCARHCGMPGTQCVNCSQFLRQEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPRECQPOKGSVTCFGEADOCVACHYADPPCVAR 600
 DB 541 VEECRVLOGLPREYVNAHCLPCHPRECQPOKGSVTCFGEADOCVACHYADPPCVAR 600
 QY 601 PGNFTVFWLVRPVASASHE---PDEGACQPCPINCSTHSCVDLDDKQCPAQRASP 656
 DB 601 PGNFTVFWLVRPVASASHE---PDEGACQPCPINCSTHSCVDLDDKQCPAQRASP 650
 QY 657 LTSIVSAVGLLVVVGVLGKIRKROQKIRKTYMRLLQETELVEPLTPSGAMPNQA 716
 DB 651 LTSIVSAVGLLVVVGVLGKIRKROQKIRKTYMRLLQETELVEPLTPSGAMPNQA 710
 QY 717 QMRILKETELRKVKVLSGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDE 776

Db 711 QMRILKETEIRKVKVLSGSAFGVYKGIPIPDGNNKIPVAIVLENTSPKANKELDE 770
 Qy 777 AYMAGVSPYVSRLLIGICTSTVOLVQLMPYGCLLDHYENRGRSGSODLLNMCQIA 836
 Db 771 AYMAGVSPYVSRLLIGICTSTVOLVQLMPYGCLLDHYENRGRSGSODLLNMCQIA 830
 Qy 837 KGMSTYEDVLYHRDLAARVLYKSPNHVKITDPGLARLDIDETEVHADGKVPIKMA 896
 Db 831 KGMSTYEDVLYHRDLAARVLYKSPNHVKITDPGLARLDIDETEVHADGKVPIKMA 890
 Qy 897 LESILRRRFTHOSDWSYGTWELMTFGAKPYDGIIPAREIPLLKGBELPQPICTID 956
 Db 891 LESILRRRFTHOSDWSYGTWELMTFGAKPYDGIIPAREIPLLKGBELPQPICTID 950
 Qy 957 VYIMVYKCMWIDSECRPRFRELVSFESRMARDQRFVYIQLNEDLGAPSLDSTFYRSLLE 1016
 Db 951 VYIMVYKCMWIDSECRPRFRELVSFESRMARDQRFVYIQLNEDLGAPSLDSTFYRSLLE 1010
 Qy 1017 DDMGDLVDAEEXLYLPQGGFFCPDPAFGAGMTHHRSSSTSGGGLTLGLPSEEEA 1076
 Db 1011 DDMGDLVDAEEXLYLPQGGFFCPDPAFGAGMTHHRSSSTSGGGLTLGLPSEEEA 1070
 Qy 1077 PRSPLAPSEAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDPVLPSETDGYAP 1136
 Db 1071 PRSPLAPSEAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDPVLPSETDGYAP 1130
 Qy 1137 LTGSPQBEYVNOVDYRPPSPREGLPAARPAATLERAKTISPKNGVYKDVAFAGGA 1196
 Db 1131 LTGSPQBEYVNOVDYRPPSPREGLPAARPAATLERAKTISPKNGVYKDVAFAGGA 1190
 Qy 1197 VENPEYLTPOGGAAPRPPAPSPAFDNLXYWDOPPRGAPSPSTFKGPTAENPEYLG 1256
 Db 1191 VENPEYLTPOGGAAPRPPAPSPAFDNLXYWDOPPRGAPSPSTFKGPTAENPEYLG 1250
 Qy 1257 LDVPV 1261
 Db 1251 LDVPV 1255
 RESULT 4
 AAU74545
 ID AAU74545 standard, Protein, 1255 AA.
 AC AAU74545;
 XX 23-APR-2002 (first entry)
 DE Human HER2 (ErbB2) polypeptide.
 XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW gliad disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastococelic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX Homo sapiens.
 OS Homo sapiens.
 PN US2002001587-A1.
 XX 03-JAN-2002.
 XX 16-MAR-2001; 2001US-0811123.
 PF 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIM/) SLIKOWSKI M.
 XX

PI Erickson S, Schwall R, Slikowski M;
 XX WPI; 2002-163686/21.
 DR N-PSDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal
 PS
 XX
 XX Example 3; Fig 7; 93pp; English.
 CC The invention relates to treating a tumour in a mammal, where the tumour
 CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastococelic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 XX
 Query Match 96.9%; Score 6627; DB 23; Length 1255;
 Best Local Similarity 97.0%; Pred. No. 0;
 Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;
 Qy 1 MELAALCRWGLLALLPPGAASVQCTGTDKLRLPASPEHLDMLRLYQCCQVQVQNL 60
 Db 1 MELAALCRWGLLALLPPGAASVQCTGTDKLRLPASPEHLDMLRLYQCCQVQVQNL 60
 Qy 61 ELTYLPTNASLSPLODIOEVGVYLAHNOVROPLORLIVRGTOLEFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSPLODIOEVGVYLAHNOVROPLORLIVRGTOLEFEDNYALAVLDNG 120
 Qy 121 DPLNNTTPVTGASPGGLRELQRLSTELKGVLIQRNPQLCYODTILMKDIFHNKQLA 180
 Db 121 DPLNNTTPVTGASPGGLRELQRLSTELKGVLIQRNPQLCYODTILMKDIFHNKQLA 180
 Qy 181 LTLIDTNSRACHPQSPKCKSRGWBSSEDCQSLRTVCAAGCARCKGPIPTDCHEQC 240
 Db 181 LTLIDTNSRACHPQSPKCKSRGWBSSEDCQSLRTVCAAGCARCKGPIPTDCHEQC 240
 Qy 241 AAGCTGPRGSDCLAHFNHSGICELCPALVTYNTTFESMPREGYTGAACVYACP 300
 Db 241 AAGCTGPRGSDCLAHFNHSGICELCPALVTYNTTFESMPREGYTGAACVYACP 300
 Qy 301 YNYLSTVGSCTVLCPLHNOEVTAEADGTORCEKSKCARVYCYGLGMOYIKANSKFIGIT 360
 Db 301 YNYLSTVGSCTVLCPLHNOEVTAEADGTORCEKSKCARVYCYGLGMEHLEREVAVYSAN 360
 Qy 361 ELEPAGCKKIKGSLAFLPESFEDGPASNTAPLOEQLOVEFTLEITGYLYISAMPDPLP 420
 Db 361 IOEPAGCKKIKGSLAFLPESFEDGPASNTAPLOEQLOVEFTLEITGYLYISAMPDPLP 420
 Qy 421 DLSVFONLOVIRGRILHNGAVSLTLOGLIGISWLGRLRLRELSGSLALIHNNTHLCFYHTV 480
 Db 421 DLSVFONLOVIRGRILHNGAVSLTLOGLIGISWLGRLRLRELSGSLALIHNNTHLCFYHTV 480
 Qy 481 PMDOLFNRPHQALHTANRPEDEVGSGGLACHQLCARGHCKGPGPTCCVNCSCPLRQEC 540
 Db 481 PMDOLFNRPHQALHTANRPEDEVGSGGLACHQLCARGHCKGPGPTCCVNCSCPLRQEC 540
 Qy 541 VEECRVYQGLPREYVYARHCLPCHPECOPONGSVTCGPEADOCVCAHYKDPFCVARC 600
 Db 541 VEECRVYQGLPREYVYARHCLPCHPECOPONGSVTCGPEADOCVCAHYKDPFCVARC 600
 Qy 601 PFNNFTVSWLRVPKVSASHLE---DDEEGACQPCPINCSTHSCVDLDDKGCAPABORASP 656

```

Db 601 PSC-----VKEDLSYMPIMKPFDEEGACQPCPCINTHSCVLDLKGCAEGRASP 650
Qy 657 LTIIVSAVVGILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPIITPSGAMPNA 716
Db 651 LTIIVSAVVGILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPIITPSGAMPNA 710
Qy 717 QMEILKETELRKVVGSSAFGTVYVGIWIPDEENKIPAIIVLEENISPKXNKILDE 776
Db 711 QMEILKETELRKVVGSSAFGTVYVGIWIPDEENKIPAIIVLEENISPKXNKILDE 770
Qy 777 AYMAGVGSFVSRLLGLCTSTVOLTQMLPYGCLLDHYRENRGLSGODLIMCMQIA 836
Db 771 AYMAGVGSFVSRLLGLCTSTVOLTQMLPYGCLLDHYRENRGLSGODLIMCMQIA 830
Qy 837 KGMSTYLEDVRLNHRDLAARVLYKSPNHVKITDFGLARLDIDETEHADGGKVPKMA 896
Db 831 KGMSTYLEDVRLNHRDLAARVLYKSPNHVKITDFGLARLDIDETEHADGGKVPKMA 890
Qy 897 LESLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTID 956
Db 891 LESLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTID 950
Qy 957 VYMIWKCMITDSECRPFRELVSSESRMARDPQRFVYIQNEDLGSPASPLDSTFYRSLLE 1016
Db 951 VYMIWKCMITDSECRPFRELVSSESRMARDPQRFVYIQNEDLGSPASPLDSTFYRSLLE 1010
Qy 1017 DDMGMDLVDAEYVLPQGGFFCPDPAPAGAGMHHRRSSSTSGGGDLTLIGPSEEEA 1076
Db 1011 DDMGMDLVDAEYVLPQGGFFCPDPAPAGAGMHHRRSSSTSGGGDLTLIGPSEEEA 1070
Qy 1077 PRSLAPSEAGSDVDFDGLGMAKGLQSLPTHDSPLQRYSEDPVPLPSETDGYVAP 1136
Db 1071 PRSLAPSEAGSDVDFDGLGMAKGLQSLPTHDSPLQRYSEDPVPLPSETDGYVAP 1130
Qy 1137 LTGSPQPEYVNOQDVRRQPSRREGPLPARAPAGATLEAKTISPGKNGVYKQVAFGGA 1196
Db 1131 LTGSPQPEYVNOQDVRRQPSRREGPLPARAPAGATLEAKTISPGKNGVYKQVAFGGA 1190
Qy 1197 VENEYVLTPOGGAAPQPPAPSPAFNDLYWDDPPEEGAPSTFKGPTAENEYVIG 1256
Db 1191 VENEYVLTPOGGAAPQPPAPSPAFNDLYWDDPPEEGAPSTFKGPTAENEYVIG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX
XX AAW01111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX breast cancer; ovary cancer; colon cancer; lung cancer;
XX prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 676..1255
FT Domain /label= Intracellular domain
FT /note="Claimed domain, useful for immunisation"
XX
PN MO630514-A1.
XX
XX 03-OCT-1996.
XX

```

```

PF 28-MAR-1996; 96MO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX MPI. 1996-455361/45.
XX
XX N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly(peptide)s - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.
XX
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
XX the product of the HER-2/neu oncogene (see also AAT40739). The
XX protein is over-expressed in various cancers, including breast,
XX ovarian, colon, lung and prostate. The intracellular domain of the
XX protein can be used to immunise an animal against a malignancy with
XX which the oncogene is associated. The polypeptide can be produced
XX in transformed host cells for use in immunisation. Alternatively,
XX animal cells are transfected in vivo or ex vivo with a viral vector
XX that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
XX
Query Match 96.8%; Score 6621; DB 17; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
Qy 1 METALCWGILLALPPGASTQVCTGDMKLRPASPEHMLRLVGGCCVQVQNL 60
Db 1 METALCWGILLALPPGASTQVCTGDMKLRPASPEHMLRLVGGCCVQVQNL 60
Qy 61 ETTYLPTNASLFLDIOEVQGYVLIANQVQVPLQRLIVRGTLFEDNYALAVLDNG 120
Db 61 ETTYLPTNASLFLDIOEVQGYVLIANQVQVPLQRLIVRGTLFEDNYALAVLDNG 120
Qy 121 PLNNTTPTVGTASPGGLRELQRLSLTELKGVLIQRNPOLCYDTILMKDIFHNQDLA 180
Db 121 PLNNTTPTVGTASPGGLRELQRLSLTELKGVLIQRNPOLCYDTILMKDIFHNQDLA 180
Qy 181 LTLIDNRSRACHPSPCKSGRCMGESSEDCQSLTRVCAAGCARCGPLPTDCHEQC 240
Db 181 LTLIDNRSRACHPSPCKSGRCMGESSEDCQSLTRVCAAGCARCGPLPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACLFENHSGICELHCPALVTYNTDIFESMPNDEGRYTGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFENHSGICELHCPALVTYNTDIFESMPNDEGRYTGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVPLNQEVTAEQDQRECKSKPCARVYCYGLMGYITANSKFIGIT 360
Db 301 YNYLSTDVSGCTLVPLNQEVTAEQDQRECKSKPCARVYCYGLMGYITANSKFIGIT 360
Qy 361 ELEFAGCKKIFGSLAFLESPFGDPASNTALQPEQLOVFETLEITGYLISAMPDLSLP 420
Db 361 IGFAGCKKIFGSLAFLESPFGDPASNTALQPEQLOVFETLEITGYLISAMPDLSLP 420
Qy 421 DLSVFQNTQVTRGRIANGAYSLLTQGLGISWLGIRSLRELGSGIALIHNTHLCEVHTV 480
Db 421 DLSVFQNTQVTRGRIANGAYSLLTQGLGISWLGIRSLRELGSGIALIHNTHLCEVHTV 480
Qy 481 PMDOLFRRPHQALLHTARPEDECEGGLACHOICARHCGMPGTQCVNCSQFLRGEC 540
Db 481 PMDOLFRRPHQALLHTARPEDECEGGLACHOICARHCGMPGTQCVNCSQFLRGEC 540
Qy 541 VEECRVLTQGLPREYVNAHCHLPCHPECCQPNQGSVTCFQPEADQCVACAHYDPPFCVARC 600
Db 541 VEECRVLTQGLPREYVNAHCHLPCHPECCQPNQGSVTCFQPEADQCVACAHYDPPFCVARC 600

```

QY	601	PPNNFTVSVFWLNVPRVYSASHLE----	PPBEAGCQPCPLNCHSSCYVDLDKGPACQRA	656
Db	601	PGG-----VKPDUSYMPIMKFP	PDEBGACQPCPLNCHSSCYVDLDKGPACQRA	650
QY	657	LTSIVSAVGLLLVVLGVVFGILIKRROQK	IRKXTMRLLQETELVEPLTPSGAMPQA	716
Db	651	LTSIIISAVGILLVVLGVVFGILIKRROQK	IRKXTMRLLQETELVEPLTPSGAMPQA	710
QY	717	QMRLLKETELRKVKVLGSGAFCTYKGI	IPGENVKIPVAIKVLRENTSPKANKEILDE	776
Db	711	QMRLLKETELRKVKVLGSGAFCTYKGI	IPGENVKIPVAIKVLRENTSPKANKEILDE	770
QY	777	AYVMAGVGSPPYVSRLIGICTSTVOVL	QLMAYGCGLLDHVRENGRISQDILNMCQIA	836
Db	771	AYVMAGVGSPPYVSRLIGICTSTVOVL	QLMAYGCGLLDHVRENGRISQDILNMCQIA	830
QY	837	KCMSYLEDVRLVHRDLAARNVLVKS	PNHYKITDFGLARLLDIDETEHADGKVP	IKXMA 896
Db	831	KCMSYLEDVRLVHRDLAARNVLVKS	PNHYKITDFGLARLLDIDETEHADGKVP	IKXMA 890
QY	897	LESILRRPETHOSDWSYGVYWE	LMTFGAKXPYDGIIPAREIPDLLEKGERLP	POPICIID 956
Db	891	LESILRRPETHOSDWSYGVYWE	LMTFGAKXPYDGIIPAREIPDLLEKGERLP	POPICIID 950
QY	957	YVMIMWKCMMIDSECRPREREL	YSEFSMARDDPQRFVYIQNEDLGPASPLD	STFYRSILE 1016
Db	951	YVMIMWKCMMIDSECRPREREL	YSEFSMARDDPQRFVYIQNEDLGPASPLD	STFYRSILE 1010
QY	1017	DDMDMDLVADEEYLVLPQCGF	PCPPDPAFGAGVHHRHSSSTRSGGDL	TLGLEPSEBEA 1076
Db	1011	DDMDMDLVADEEYLVLPQCGF	PCPPDPAFGAGVHHRHSSSTRSGGDL	TLGLEPSEBEA 1070
QY	1077	PRSPPLAPSEAGSDVDFDGLG	WAKGIQSLPTHDPSPLOKYSDEPTVPL	LPSETGTGYAP 1136
Db	1071	PRSPPLAPSEAGSDVDFDGLG	WAKGIQSLPTHDPSPLOKYSDEPTVPL	LPSETGTGYAP 1130
QY	1137	LTCSPQPEYVNOPTVRPOP	PSRREBPILPAAPACATLERPXTLS	SGKXGVKDVAFGGA 1196
Db	1131	LTCSPQPEYVNOPTVRPOP	PSRREBPILPAAPACATLERPXTLS	SGKXGVKDVAFGGA 1190
QY	1197	VENPEYLVLPQCGAQP	HPHPAPFSAFNLVYVDDP	PERGAPSTFPGTETANPEYIG 1256
Db	1191	VENPEYLVLPQCGAQP	HPHPAPFSAFNLVYVDDP	PERGAPSTFPGTETANPEYIG 1250
QY	1257	LDVPY 1261		
Db	1251	LDVPY 1255		
RESULT 6				
AAW92406				
ID	AAW92406 standard; Protein; 1255 AA.			
XX	AAW92406;			
XX	DT 21-APR-1999 (first entry)			
DE	Human HER-2/neu oncogene protein.			
KW	HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;			
XX	malignancy; treatment; tumour.			
XX	Homo sapiens.			
FT	Key	Location/Qualifiers		
FT	Region	676..1255		
FT		/note= "region which elicits immune response"		
XX	US5869445-A.			
XX	09-FEB-1999.			
XX	01-APR-1996;	96US-0625101.		

QY 601 PFNNFTVSWLRVPKVASHLE----PDEGACQPCPINCSTHSCVDLDKGCAGPABSP 656
 DB 601 PSQ-----VKPDLSTYMPIMKFPDEGACQPCPINCSTHSCVDLDKGCAGPABSP 650
 QY 657 LSTISAVVGLLVVVGAVVFGILIKRQOKIRKYTRRLLOETELVEPLTPSGAPNQA 716
 DB 651 LSTISAVVGLLVVVGAVVFGILIKRQOKIRKYTRRLLOETELVEPLTPSGAPNQA 710
 QY 717 QMRILKETELRKVKVUGSAGPVGIVYKGIWIPDGENVKI PAIVKLRNTSPKANKELIDE 776
 DB 711 QMRILKETELRKVKVUGSAGPVGIVYKGIWIPDGENVKI PAIVKLRNTSPKANKELIDE 770
 QY 777 AVYVAGVSPVYVSLILICTSTVQVLTQMPYQCLLDHRENRGRAGSDILNMCQQA 836
 DB 771 AVYVAGVSPVYVSLILICTSTVQVLTQMPYQCLLDHRENRGRAGSDILNMCQQA 830
 QY 837 KGMGYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 896
 DB 831 KGMGYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 890
 QY 897 LESTLRRTQSGDVMSVGYVMEMLTFGAKPVYGI PAREIPDLLEKGERLPQPICCTID 956
 DB 891 LESTLRRTQSGDVMSVGYVMEMLTFGAKPVYGI PAREIPDLLEKGERLPQPICCTID 950
 QY 957 VYMIWVKCMMIDSECRPFRELYSEFSRMAPDPQRFVVIQNEIDGSPASPLDSTFYSLLE 1016
 DB 951 VYMIWVKCMMIDSECRPFRELYSEFSRMAPDPQRFVVIQNEIDGSPASPLDSTFYSLLE 1010
 QY 1017 DDDMGDLVDAEYLYVPOQGFPCPDPAAGAGMTHRRSSSTSGGDLTLGLEPSEEA 1076
 DB 1011 DDDMGDLVDAEYLYVPOQGFPCPDPAAGAGMTHRRSSSTSGGDLTLGLEPSEEA 1070
 QY 1077 PRSLPASEGAGSDVFGDGLGMAKGLQSLPTHPDPLQRYSEDPVPLPSETDGYVAP 1136
 DB 1071 PRSLPASEGAGSDVFGDGLGMAKGLQSLPTHPDPLQRYSEDPVPLPSETDGYVAP 1130
 QY 1137 LTCSPQPEYVNPQPVPRQPPSPREGPLPAAPAGATLERKXTLSPGKGVYKQVFAFGA 1196
 DB 1131 LTCSPQPEYVNPQPVPRQPPSPREGPLPAAPAGATLERKXTLSPGKGVYKQVFAFGA 1190
 QY 1197 VENPEYLTPOGGAPOHPPPAPAFNDLYVMQDPPREGAPSTFKGTPTAENPEYLG 1256
 DB 1191 VENPEYLTPOGGAPOHPPPAPAFNDLYVMQDPPREGAPSTFKGTPTAENPEYLG 1250
 QY 1257 LDVBPV 1261
 DB 1251 LDVBPV 1255

RESULT 7
 AAB21198
 ID AAB21198 standard; protein; 1255 AA.

XX AAB21198;
 AC AAB21198;
 XX 12-JAN-2001 (first entry)
 DT 12-JAN-2001 (first entry)
 DE Human HER-2/neu protein.
 XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 XX colon cancer.
 OS Homo sapiens.
 XX WO200044899-A1.
 XX 03-AUG-2000.
 XX 28-JAN-2000; 2000WO-US02164.
 XX 29-JAN-1999; 99US-0117976.
 XX

PA (CORI-) CORIYA CORP.
 PA (SMK) SMITHKLINE BEECHAM.
 PI Cheever MA, Gheyssen D;
 XX MPI; 2000-505976/45.
 DR N-PsDB; AAA89736.
 XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PS prostate cancers -
 PS Claim 52; Fig 7; 128pp; English.
 CC The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against
 CC these neoplasias.
 CC
 SQ Sequence 1255 AA;
 Query Match 96.8%; Score 6621; DB 21; Length 1255;
 Best Local Similarity 96.8%; Pred. No. 0;
 Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
 QY 1 MELALCRWGLLALLPQGAATGYCTGTDMKLRPASBETHLDMRLHYOGCQVAGNTL 60
 DB 1 MELALCRWGLLALLPQGAATGYCTGTDMKLRPASBETHLDMRLHYOGCQVAGNTL 60
 QY 61 ELTYLPTNASLSFLDIOEVQGYVLIANOVROVPLQRLRIYRGTLQFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLSFLDIOEVQGYVLIANOVROVPLQRLRIYRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVTSAGPGLRLQRLSLTEILKGVLLQRPOLCYODTILMKDFHKNOLA 180
 DB 121 DPLNNTPTVTSAGPGLRLQRLSLTEILKGVLLQRPOLCYODTILMKDFHKNOLA 180
 QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTYCAGGACARCKPLPTDCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTYCAGGACARCKPLPTDCHEQC 240
 QY 241 AAGCTGPRKSDCLACHFNHSGICEHCPALVTYNTDFESMNPREGYTFEGASCVTACP 300
 DB 241 AAGCTGPRKSDCLACHFNHSGICEHCPALVTYNTDFESMNPREGYTFEGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGTQRCCKSKPCARVYGLGMOYIKANSKFIGIT 360
 DB 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGTQRCCKSKPCARVYGLGMOYIKANSKFIGIT 360
 QY 361 ELDFAGCKKI FGSLLAPLPSSFDGDPASNTAPLQPOLOVFTLEETITGLYISAMPDSLP 420
 DB 361 IOEFAGCKKI FGSLLAPLPSSFDGDPASNTAPLQPOLOVFTLEETITGLYISAMPDSLP 420
 QY 421 DLSVFONLQVIRGRIHNGAYSLTLOGLISWLGRLSRLRELSGGLALIHNNHLCFVHTV 480
 DB 421 DLSVFONLQVIRGRIHNGAYSLTLOGLISWLGRLSRLRELSGGLALIHNNHLCFVHTV 480
 QY 481 PMDOLFRNPHQALLHTANPEDECVGEGLACHQLCARGHGWPSPGTCCNCSQFLRGDEC 540
 DB 481 PMDOLFRNPHQALLHTANPEDECVGEGLACHQLCARGHGWPSPGTCCNCSQFLRGDEC 540
 QY 541 VEECRVLOGLPREYVNAARCLPCHEPCQONGSVTCFGEPAQCYACAHYKPPPCVARC 600
 DB 541 VEECRVLOGLPREYVNAARCLPCHEPCQONGSVTCFGEPAQCYACAHYKPPPCVARC 600
 QY 601 PFNNFTVSWLRVPKVASHLE----PDEGACQPCPINCSTHSCVDLDKGCAGPABSP 656

```

Db 601 PSG-----VKPDLSTYPMWKFPDEEGACOPCPINCHSCVDLDDKGPAPQASR 650
QY 657 LTSIVSAVGLILVYVGVVGLIKRQOKIRKRYTRRLLOETELVEPIBPGAMPNQA 716
Db 651 LTIISAVVILLVVGVVGLIKRQOKIRKRYTRRLLOETELVEPIBPGAMPNQA 710
QY 717 QMBILKTELKRVKVLGSGAFGVYKGIWIPDENVKI PAIVYLENTSPKANKILDE 776
Db 711 QMBILKTELKRVKVLGSGAFGVYKGIWIPDENVKI PAIVYLENTSPKANKILDE 770
QY 777 AYVWAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDHYENRGRIGSODLLNMCQIA 836
Db 771 AYVWAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDHYENRGRIGSODLLNMCQIA 830
QY 837 KGNMYSLEDAVLVHRLDAARVLYKSPNHVKITDPGLARLIDIDETEXHADGKVPKMA 896
Db 831 KGNMYSLEDAVLVHRLDAARVLYKSPNHVKITDPGLARLIDIDETEXHADGKVPKMA 890
QY 897 LESLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKGERLPQPICTID 956
Db 891 LESLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKGERLPQPICTID 950
QY 957 VVWIMVCMWIDSECRFRELVSSEPRMARDPQRFVVIQNEPLGSPASPLDSTFYSLLE 1016
Db 951 VVWIMVCMWIDSECRFRELVSSEPRMARDPQRFVVIQNEPLGSPASPLDSTFYSLLE 1010
QY 1017 DDDMGDLVDAEYVLPQGFCCDPAPAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYVLPQGFCCDPAPAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1070
QY 1077 PRSPPLABEGAGSDVFPDGLGMAKAGLQSLPTHDPSPLORYSEDPVLPSPSTDGYAP 1136
Db 1071 PRSPPLABEGAGSDVFPDGLGMAKAGLQSLPTHDPSPLORYSEDPVLPSPSTDGYAP 1130
QY 1137 LTCSPQPEYVNOQDVRPQPSPREGPPAPAPAGATLERAKTLSPPKNGVWQVAFPGA 1196
Db 1131 LTCSPQPEYVNOQDVRPQPSPREGPPAPAPAGATLERAKTLSPPKNGVWQVAFPGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPAPSPADNLVYWDODPPERGAPSTFKGTFTANDEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPAPSPADNLVYWDODPPERGAPSTFKGTFTANDEYLG 1250
QY 1257 LDVPPV 1261
Db 1251 LDVPPV 1255

```

RESULT 8
 AAY84780
 ID AAY84780 standard; Protein: 1255 AA.

XX AAY84780;
 DT 08-AUG-2000 (first entry)
 XX
 DE Amino acid sequence of the SPICE erbB-2 receptor protein.
 XX
 KW SPICE erbB-2 receptor protein; cell transformation disorder; cancer;
 KW tumor cell proliferation; tissue degeneration; arthropathy;
 KW bone resorption; inflammatory disease; degenerative disorder;
 KW wound healing.
 OS Homo sapiens.
 XX
 XX MO2000020579-A1.
 XX
 PD 13-APR-2000.
 XX
 PF 01-OCT-1999; 99MO-CA00912.
 XX
 PR 02-OCT-1998; 98US-0165192.
 XX

```

PA (UYN-) UNIV MCMMASTER.
XX
PI Muller WJ, Siegel PM;
XX WPI, 2000-103768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPICE
XX erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3, Fig 2, 60pp, English.
XX
CC The present sequence represents a SPICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;
Query Match 96.8%; Score 6621; DB 21; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
QY 1 MELALCRWGLLALLPPGAATVCTGTDMKLRLPASPTLMDLRLHYOGQVYQGNL 60
Db 1 MELALCRWGLLALLPPGAATVCTGTDMKLRLPASPTLMDLRLHYOGQVYQGNL 60
QY 61 ELTYVPTNASSFLDIOEVQGYVLAHQVQVPLQSLRLVGTQLFEDNVYALAVDNG 120
Db 61 ELTYVPTNASSFLDIOEVQGYVLAHQVQVPLQSLRLVGTQLFEDNVYALAVDNG 120
QY 121 PLNNTTPTVGTASPGGLRELQRLSTELKGVLIQRNPOLCYOTIIMKQIFHNQOLA 180
Db 121 PLNNTTPTVGTASPGGLRELQRLSTELKGVLIQRNPOLCYOTIIMKQIFHNQOLA 180
QY 122 PLNNTTPTVGTASPGGLRELQRLSTELKGVLIQRNPOLCYOTIIMKQIFHNQOLA 180
Db 122 PLNNTTPTVGTASPGGLRELQRLSTELKGVLIQRNPOLCYOTIIMKQIFHNQOLA 180
QY 181 LTLIDNRSRAHPSPCKSGSRGESSSDQSLTRTVCAAGCARCKPLPTDCHSQC 240
Db 181 LTLIDNRSRAHPSPCKSGSRGESSSDQSLTRTVCAAGCARCKPLPTDCHSQC 240
QY 241 AAGCTGPRGSDCLAFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPRGSDCLAFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTVCLPCLNNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLEVRATVSAN 360
Db 301 YNYLSTDVSGCTVCLPCLNNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLEVRATVSAN 360
QY 361 ELFPAGCKKIFGSLAFPESEFDGPASNTAPLQEPOLQVFETLEITGYLTISAMPDLP 420
Db 361 IQEPAGCKKIFGSLAFPESEFDGPASNTAPLQEPOLQVFETLEITGYLTISAMPDLP 420
QY 421 DLSVFNQNLQVIRGRIIHNGAVSLTLOGIGISLGLRSRLREGSGALAIHNTHTLCFVHTV 480
Db 421 DLSVFNQNLQVIRGRIIHNGAVSLTLOGIGISLGLRSRLREGSGALAIHNTHTLCFVHTV 480
QY 481 PMDOLFENPHQALLHTANRPEDECVGEIACQLCARGHCWGPPTOCVNSQPLRGQEC 540
Db 481 PMDOLFENPHQALLHTANRPEDECVGEIACQLCARGHCWGPPTOCVNSQPLRGQEC 540
QY 541 VEECRVTLQGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQCVAACAHYKDPFVYARC 600
Db 541 VEECRVTLQGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQCVAACAHYKDPFVYARC 600
QY 601 PFNNFTVSWFLRVKVASHLE----PDSEGAQCPCTINCHSCVDLDDKGPAPQASR 656

```

```

Db 601 PSG-----VKPDLSTYMPITWKFPPDEGACQPCPCHTCHSCVDLDDGCPAEQPSAD 650
Qy LTSIVSAVGLLVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQ 716
Db 651 LTSIVSAVGLLVVVLGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQ 710
Qy QMRILKETELRKVVLGSSAFGTYVKIWIIPDGENVKIPIAIVLRENTSPKANKELDE 776
Db 711 QMRILKETELRKVVLGSSAFGTYVKIWIIPDGENVKIPIAIVLRENTSPKANKELDE 770
Qy AYVWAGVSEFVSSLLICITSTVQLTQCLMPYGCCLDHRNRRGLSGODLNMCMQIA 836
Db 771 AYVWAGVSEFVSSLLICITSTVQLTQCLMPYGCCLDHRNRRGLSGODLNMCMQIA 830
Qy KGMSTEDVRLVHRDLAARNVLKSPNHVKITDPGLARLIDIDETEHADGKVPFKMA 896
Db 831 KGMSTEDVRLVHRDLAARNVLKSPNHVKITDPGLARLIDIDETEHADGKVPFKMA 890
Qy LESTLRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 956
Db 891 LESTLRRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 950
Qy VYMIWVKCMMIDSECRPFRELVESESRMARDPQRFVYIQNEDLGPAAPLDSTFYRSLLE 1016
Db 951 VYMIWVKCMMIDSECRPFRELVESESRMARDPQRFVYIQNEDLGPAAPLDSTFYRSLLE 1010
Qy DDDMGDLVDAEYVLPQGFPCPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYVLPQGFPCPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1070
Qy PRSPPLASEGAGSDVFPDGLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPPLASEGAGSDVFPDGLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAP 1130
Qy LTGSPQPEYVNPQPVPRQPPSPREGPLPAAPPAATIERAKTISPGKNGYVVKQVFAFGA 1196
Db 1131 LTGSPQPEYVNPQPVPRQPPSPREGPLPAAPPAATIERAKTISPGKNGYVVKQVFAFGA 1190
Qy VENPEYLTPOGGAPOHPHPAPAFSPAFDNLYYMDQDPPERGAAPSTPKGTANPEYVIG 1256
Db 1191 VENPEYLTPOGGAPOHPHPAPAFSPAFDNLYYMDQDPPERGAAPSTPKGTANPEYVIG 1250
Qy LDVEPV 1261
Db 1251 LDVEPV 1255

```

RESULT 9
AAB85458 standard; Protein; 1255 AA.

XX AAB85458;
XX 25-SEP-2001 (first entry)
XX Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX 26-JUL-2001.
XX 19-JAN-2001; 2001WO-US01850.
XX 21-JAN-2000; 2000US-0177545.
XX (CORI-) CORIXA CORP.

```

PI Cheever MA, Hand-Zimmermann S;
XX WPI; 2001-476112/51.
DR N-PSDB; AAB23392.
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
for treating or preventing cancer. e.g. breast cancer
XX Claim 2; Page 41-46; 49pp; English.
XX The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer. e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX Sequence 1255 AA;
XX
Query Match 96.8%; Score 6621; DB 22; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

```

Db 651 LSTIISAVGILLVVLGVVFGIILIKRQOKIRKTYNRLLQETELVEPLTPSGAMENQA 710
 QY 717 QMRILKETELRKXVVLGSGAFGTVYKGIWIPDGENVKIPAIKVLRENTSPKANKETLDE 776
 Db 711 QMRILKETELRKXVVLGSGAFGTVYKGIWIPDGENVKIPAIKVLRENTSPKANKETLDE 770
 QY 777 AYVWAGVGSFVYSRLGICITSTVQLVTLQMPYGCCLLDHVENRGRIGSODLLNMCQIA 836
 Db 771 AYVWAGVGSFVYSRLGICITSTVQLVTLQMPYGCCLLDHVENRGRIGSODLLNMCQIA 830
 QY 837 KGMVLEDEVLRHDLAARVVLKSPVHVKITDGLARLLDIDETEHADGGVPIKMA 896
 Db 831 KGMVLEDEVLRHDLAARVVLKSPVHVKITDGLARLLDIDETEHADGGVPIKMA 890
 QY 897 LESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 956
 Db 891 LESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 950
 QY 957 VYIMVWCMIMIDECRRFRRELVSFSRMARDQRFVYIQNEDLGSPASPLDSTFYRSLLE 1016
 Db 951 VYIMVWCMIMIDECRRFRRELVSFSRMARDQRFVYIQNEDLGSPASPLDSTFYRSLLE 1010
 QY 1017 DDMGDLVDAEEYLVPOQGFCCDPAPAGAGMHHRRSSSTRSGGDLTLGSPSEEEA 1076
 Db 1011 DDMGDLVDAEEYLVPOQGFCCDPAPAGAGMHHRRSSSTRSGGDLTLGSPSEEEA 1070
 QY 1077 PRSPPLABEGAGSDVDPDGLQMGNAKGLSLPHDPSLPQRYSEDEPTVPLPSETDGYVAP 1136
 Db 1071 PRSPPLABEGAGSDVDPDGLQMGNAKGLSLPHDPSLPQRYSEDEPTVPLPSETDGYVAP 1130
 QY 1137 LTCSPOPEYVNOPDVRQPPSPREGPLPARPAGATLERAKTLSPKNGVWVAFAGCA 1196
 Db 1131 LTCSPOPEYVNOPDVRQPPSPREGPLPARPAGATLERAKTLSPKNGVWVAFAGCA 1190
 QY 1197 VENPEYLTPOGGAAPQHPHPAPSPAFDNLVYWDODPBERGAPSTFFKGTPTANPEYLG 1256
 Db 1191 VENPEYLTPOGGAAPQHPHPAPSPAFDNLVYWDODPBERGAPSTFFKGTPTANPEYLG 1250
 QY 1257 LDVPV 1261
 Db 1251 LDVPV 1255
 RESULT 10
 AAG88267
 ID AAG88267 standard; Protein; 1255 AA.
 AC AAG88267;
 DT 11-SEP-2001 (first entry)
 DE HER2/neu amino acid sequence.
 KM Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KM Immune response; vaccine; tumor; cancer; cytotoxic; immunostimulant;
 KM Tumor-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
 OS Homo sapiens.
 PN MO200141787-A1.
 PD 14-JUN-2001.
 PF 11-DEC-2000; 2000MO-US33591.
 PR 10-DEC-1999; 99US-0458299.
 PA (EPIM-) EPIMUNE INC.
 PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E,
 PI Keogh E;
 XX MPI, 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer.
 XX
 PS Disclosure, Page 15, 1999p, English.

CC The present invention describes isolated prepared HER2/neu epitopes (I).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (II), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic
 CC and immunostimulant activities, and can be used in vaccines (I), (II)
 CC and (III) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (I) and (II) are useful for
 CC monitoring or evaluating an immune response to a tumor-associated
 CC antigen when incubated with a T lymphocyte sample from a patient and
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumor
 CC vaccine provides the opportunity to combine epitopes derived from
 CC multiple tumor-associated molecules to address the problem of tumor-
 CC tumor variability and reducing the likelihood of tumor escape due to
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
 CC the exemplification of the present invention.

Sequence 1255 AA;

Query Match 96.8%; Score 6621; DB 22; Length 1255;
 Best Local Similarity 96.8%; Pred. No. 0;
 Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

QY 1 MELALICRWGILLALPPGAASVQCTGCTDMKRLPASPETHLDMRLHYGCGVYQGNL 60
 Db 1 MELALICRWGILLALPPGAASVQCTGCTDMKRLPASPETHLDMRLHYGCGVYQGNL 60
 QY 61 ELTYLPNALSFLDIOEVGVYLAHNVQVPLQRLRIVRGTOLEEDNVALAVLDNG 120
 Db 61 ELTYLPNALSFLDIOEVGVYLAHNVQVPLQRLRIVRGTOLEEDNVALAVLDNG 120
 QY 121 DPLNNTFTVGTASFGGARELQRLSTELIKGVLIQRPOLCYODTILMKDIFHKNNQLA 180
 Db 121 DPLNNTFTVGTASFGGARELQRLSTELIKGVLIQRPOLCYODTILMKDIFHKNNQLA 180
 QY 181 LTLIDTNRSRACHPCSPKCKSGCWGESSSEDCSLTRTVACAGGACRGPPLPTDCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPKCKSGCWGESSSEDCSLTRTVACAGGACRGPPLPTDCHEQC 240
 QY 241 AAGCTGPRGSCCLACLNHNSGICELHPALVYNTDTFESMPNPEGRTYTGASCVTACP 300
 Db 241 AAGCTGPRGSCCLACLNHNSGICELHPALVYNTDTFESMPNPEGRTYTGASCVTACP 300
 QY 301 YNYLSTDVGSCTLVCPLEHNOEVTADGTORCEKSKCARCYGLGMQYIKANSKFTGIT 360
 Db 301 YNYLSTDVGSCTLVCPLEHNOEVTADGTORCEKSKCARCYGLGMQYIKANSKFTGIT 360
 QY 361 ELERAGCKKIFGSLAFPESEFGDPASNTAPLQEOQLVFETLEITGYLYISAMPDLP 420
 Db 361 ELERAGCKKIFGSLAFPESEFGDPASNTAPLQEOQLVFETLEITGYLYISAMPDLP 420
 QY 421 DLSVFNQVYRGRILNNGAYSLTGLGLISWGLRSLRELGSGLAIHHNTHLCFVHTV 480
 Db 421 DLSVFNQVYRGRILNNGAYSLTGLGLISWGLRSLRELGSGLAIHHNTHLCFVHTV 480
 QY 481 PMDLFRNPHOALHTANRBEDECVGGLACHQLCARGHGCMGPPTCVNCSOPLRQEC 540

```

Db      481 PMDLPRNHQALLHTANRPEDECVGEGLAChOLCARGHCMGPGPTQCVCNCSQFLRGCEC 540
Qy      541 VEECRVLOGLPREVYNARHCLPCHPECOPOGNSVTCGEPADCCVACAHYKDPFCVABC 600
Db      541 VEECRVLOGLPREVYNARHCLPCHPECOPOGNSVTCGEPADCCVACAHYKDPFCVABC 600
Qy      601 PFNFNTVSPFLRVKPVKASHLE-----PDEGACOPCINCTHSCVLDKDGCAEGRASP 656
Db      601 PSC-----VKPRLSTMPIMKFPDEGACOPCINCTHSCVLDKDGCAEGRASP 650
Qy      657 LSTISAVVGLLVVVLGVVFGILIKRQCKIRKYMRLLOETLEVEPLPSGAMPNGA 716
Db      651 LSTISAVVGLLVVVLGVVFGILIKRQCKIRKYMRLLOETLEVEPLPSGAMPNGA 710
Qy      717 QMRILKTELKRVKVGSGAFGTVYKIMIPDEGNVYKIPALIVLENSPKANKILDE 776
Db      711 QMRILKTELKRVKVGSGAFGTVYKIMIPDEGNVYKIPALIVLENSPKANKILDE 770
Qy      777 AYVAGVSPVSRLLGICLTSTVQLTQIMPYGCLLDHYENRGRGLSGQDLINMGQIA 836
Db      771 AYVAGVSPVSRLLGICLTSTVQLTQIMPYGCLLDHYENRGRGLSGQDLINMGQIA 830
Qy      837 KGSMSYLEDVRLVRBDLAARVLYKSPNHVKITDPGLARLIDETETVHADGKVPKMA 896
Db      831 KGSMSYLEDVRLVRBDLAARVLYKSPNHVKITDPGLARLIDETETVHADGKVPKMA 890
Qy      897 LESTLRRTFHOSDVMISYGVTVMLTFGAKPYDGIPAREIPDLEKGERLPQPICTID 956
Db      891 LESTLRRTFHOSDVMISYGVTVMLTFGAKPYDGIPAREIPDLEKGERLPQPICTID 950
Qy      957 VYIMVYKMMIDSECRFRELVSFESRMARDPQRFVIVIONEDLGASPLDSTFYRSLE 1016
Db      951 VYIMVYKMMIDSECRFRELVSFESRMARDPQRFVIVIONEDLGASPLDSTFYRSLE 1010
Qy      1017 DDMGDLVDAEEXYVPOGGRFCPPAPAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1076
Db      1011 DDMGDLVDAEEXYVPOGGRFCPPAPAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1070
Qy      1077 PRSLAPSEAGSDVDFDGLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSESTDGVAP 1136
Db      1071 PRSLAPSEAGSDVDFDGLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSESTDGVAP 1130
Qy      1137 LTGSPPEYVNOQPVPRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVYKQVPAFGA 1196
Db      1131 LTGSPPEYVNOQPVPRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVYKQVPAFGA 1190
Qy      1197 VENPEXYLTPGGGAAPQHPAPAPAFDNLVYMDQDPERGAPSPSTFKGTANDEBYLG 1256
Db      1191 VENPEXYLTPGGGAAPQHPAPAPAFDNLVYMDQDPERGAPSPSTFKGTANDEBYLG 1250
Qy      1257 LDVPEV 1261
Db      1251 LDVPEV 1255

```

RESULT 11

AAE24067 standard; Protein; 1255 AA.

```

ID      AAE24067 standard; Protein; 1255 AA.
XX      AAE24067;
XX      23-SEP-2002 (first entry)
XX      Human Her-2 protein.
XX      Human Her-2 protein.
XX      Human Her-2; epidermal growth factor receptor 2; infection; cancer;
XX      hyperproliferative disorder; prophylaxis; inflammation; antisense;
XX      tumour; gene therapy; phosphorothioate backbone.
OS      Homo sapiens.
XX      W0200222636-A1.
XX

```

```

PD      21-MAR-2002.
XX      12-SEP-2001; 2001WO-US28572.
XX      15-SEP-2000; 2000US-0663834.
XX      (ISIS-) ISIS PHARM INC.
XX      Bennett CF, Coweart LM;
XX      WPI; 2002-471192/50.
XX      N-PSDB; AAD38904.
XX      Novel antisense oligonucleotide which modulates the expression of Human
XX      Epidermal Growth Factor receptor, Her2, is useful for treating tumors
XX      inflammation or to prevent infection in humans.
XX      Example 13; Page 95-107; 116pp; English.
XX      The invention relates to antisense compounds targeted to a nucleic
XX      acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX      that specifically hybridizes with and inhibits the expression of Her2.
XX      Antisense compounds of the invention are used for treating diseases or
XX      conditions associated with Her2 such as hyperproliferative disorders
XX      e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX      neural or cardiac cancer. They are also useful prophylactically e.g.
XX      to prevent or delay infection, inflammation and tumour formation. The
XX      invention is also used in gene therapy. The present sequence is human
XX      Her-2 protein.
XX      Sequence 1255 AA;
XX      Query Match 96.8%; Score 6621; DB 23; Length 1255;
XX      Best Local Similarity 96.8%; Pred. No. 0;
XX      Matches 1255; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
Qy      1 MELALCRWGLLALLPQAASTVCTGTMKRLPASPETHLDLRLHYOCQVYQNL 60
Db      1 MELALCRWGLLALLPQAASTVCTGTMKRLPASPETHLDLRLHYOCQVYQNL 60
Qy      61 ETTYPTNALSFLDIDIEVQGYVLIANNQVQVPLQRLIRYQGLEFEDNYALAVLNG 120
Db      61 ETTYPTNALSFLDIDIEVQGYVLIANNQVQVPLQRLIRYQGLEFEDNYALAVLNG 120
Qy      121 DELNNTPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYCTIIMKDIFFKNNOLA 180
Db      121 DELNNTPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYCTIIMKDIFFKNNOLA 180
Qy      121 DELNNTPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYCTIIMKDIFFKNNOLA 180
Db      121 DELNNTPTGASPGGLRELOLRSLTEILKGVLIQRPOLCYCTIIMKDIFFKNNOLA 180
Qy      181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLETDCHEQC 240
Db      181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLETDCHEQC 240
Qy      241 AAGCTGPRHSDCLACLNHNSGICELHCPALVTYNTDTFESMPNPEGVYTCASVTCAP 300
Db      241 AAGCTGPRHSDCLACLNHNSGICELHCPALVTYNTDTFESMPNPEGVYTCASVTCAP 300
Qy      301 YNYLSTDVGSCTLVCPILHNOEVTADGTQCEKSKPCARVYGLAMQYIRANSKFICIT 360
Db      301 YNYLSTDVGSCTLVCPILHNOEVTADGTQCEKSKPCARVYGLAMQYIRANSKFICIT 360
Qy      361 ELIEFAGCKKIFGSLAFLESEFDGDPASNTAPLQEPOLQVFTLEITGYLYISAMPDLP 420
Db      361 IGFAGCKKIFGSLAFLESEFDGDPASNTAPLQEPOLQVFTLEITGYLYISAMPDLP 420
Qy      421 DLSTVQNLQVIRGRLIHNGAYSLTQGGISMLGLRSRLREGSGLATLHNHTHLCFYVTV 480
Db      421 DLSTVQNLQVIRGRLIHNGAYSLTQGGISMLGLRSRLREGSGLATLHNHTHLCFYVTV 480
Qy      481 FMDQLFRNPHQALLHTANRPEDECVGEGLAChOLCARGHCMGPGPTQCVCNCSQFLRGCEC 540
Db      481 FMDQLFRNPHQALLHTANRPEDECVGEGLAChOLCARGHCMGPGPTQCVCNCSQFLRGCEC 540
Qy      541 VEECRVLOGLPREVYNARHCLPCHPECOPOGNSVTCGEPADCCVACAHYKDPFCVABC 600

```

```

Db      541 VEECRVLQGLPREYVNAHCLPCHPECOFONGSVTCFGEADQCVACAHCAPPCVAC 600
Qy      601 PENNFVFSFMRKPKYSASHLE-----PDEEGACOPCPINCTSCVDLDKGCAPAGRASP 656
Db      601 PSC-----VKPDLSTWPMKFPDEEGACQCPINCTSCVDLDKGCAPAGRASP 650
Qy      657 LRSIVSAVVGILLVVLGVFGILLIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOA 716
Db      651 LRSIVSAVVGILLVVLGVFGILLIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOA 710
Qy      717 QMRILKETELRKVKYLGSAFGVYKGIIPGGENVKIIVALKVLRNTPSPANKELIDE 776
Db      711 QMRILKETELRKVKYLGSAFGVYKGIIPGGENVKIIVALKVLRNTPSPANKELIDE 770
Qy      777 AAYMAGVSPYVSRLLIGICLTSTVQLTOLMPYGCILLDHVRENGRLSGQDILLNMQOJA 836
Db      771 AAYMAGVSPYVSRLLIGICLTSTVQLTOLMPYGCILLDHVRENGRLSGQDILLNMQOJA 830
Qy      837 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMA 896
Db      831 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMA 890
Qy      897 LESILRRPETHOSDVMYGVYWEIMTFGAKPYDGIIPAREIDPLEKGRLEQPPICITD 956
Db      891 LESILRRPETHOSDVMYGVYWEIMTFGAKPYDGIIPAREIDPLEKGRLEQPPICITD 950
Qy      957 VYMIWYKCMIDSECPREFRELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db      951 VYMIWYKCMIDSECPREFRELVSEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy      1017 DDDMGDLVAEEYLVQGGFPCPDPAFGAGVHHRHSSSTRSGGDTLTGLSESEEA 1076
Db      1011 DDDMGDLVAEEYLVQGGFPCPDPAFGAGVHHRHSSSTRSGGDTLTGLSESEEA 1070
Qy      1077 PRSPPLAPSEAGSDVDFDGLGNGAAGKGLQSLPTHPSPLQRYSEDPTVPLPSETDGYAP 1136
Db      1071 PRSPPLAPSEAGSDVDFDGLGNGAAGKGLQSLPTHPSPLQRYSEDPTVPLPSETDGYAP 1130
Qy      1133 LTRSPQPEYVNOPDVRFPPSPREGPLPAPAPAGATLEBAKTLSPGKGVVXDVAFPGA 1196
Db      1131 LTRSPQPEYVNOPDVRFPPSPREGPLPAPAPAGATLEBAKTLSPGKGVVXDVAFPGA 1190
Qy      1197 VENPEYLTPOGGAAPQPPPAFSPAFDLYYWDDPPRGAPOSTFGPTPAENPEYLG 1256
Db      1191 VENPEYLTPOGGAAPQPPPAFSPAFDLYYWDDPPRGAPOSTFGPTPAENPEYLG 1250
Qy      1257 LDVPV 1261
Db      1251 LDVPV 1255

```

```

RESULT 12
AAE20479 standard, Protein, 1255 AA.

```

```

ID      AAE20479 standard, Protein, 1255 AA.
XX      AAE20479;
XX      01-JUL-2002 (first entry)
XX      Human Her-2/neu protein.
XX      Human Her-2/neu protein.
XX      Human, Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX      human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX      Homo sapiens.
XX      Key
XX      Location/Qualifiers
XX      FT      Region
XX      /note="Naturally processed HLA-B44-restricted epitope"
XX      MO200214503-A2.

```

```

PD      21-FEB-2002.
XX      14-AUG-2001; 2001WO-US41733.
PF      14-AUG-2001; 2000US-225152P.
XX      14-AUG-2000; 2000US-225152P.
PR      28-SEP-2000; 2000US-236428P.
PR      21-FEB-2001; 2001US-270520P.
XX      (CORI-) CORIXA CORP.
PI      Hand-Zimmermann S, Cheever MA, Foy TM, Lodes NJ, Kalos MD;
PI      McNeill PD, Vedyick TS;
XX      WPI; 2002-280758/32.
DR      N-PSDB; AAD32743.
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX      prevention and diagnosis of cancer, preferably breast cancer -
XX      Disclosure; Page 114-117; 129pp; English.
XX      The invention relates to an isolated Her-2/Neu polypeptide composition
XX      effective for eliciting an immune response. The invention is useful for
XX      eliciting an immune response in a patient, where the patient is human
XX      leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX      The composition is useful for the therapy and diagnosis of cancer,
XX      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX      and other compositions for the diagnosis, prevention and treatment of
XX      human malignancies, for stimulating and/or expanding T cells specific for
XX      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX      patient. The invention is useful for stimulating a T cell response in a
XX      human patient, as probe or primer for nucleic acid hybridisation, to
XX      selectively form duplex molecules with complementary stretches of the
XX      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX      length gene from a suitable library, and to direct expression of a
XX      polypeptide in appropriate host cells. The composition is useful in
XX      prophylactic or therapeutic applications and for the treatment of cancer,
XX      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX      associated malignancies. The invention is useful in gene therapy. The
XX      present sequence is human Her-2/neu protein.
SQ      Sequence 1255 AA;
Query Match 96.8%; Score 6621; DB 23; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
Qy      1 MELAALCRMGILLALLPFGAASVQVCTGTMKRLPASPETHLDMRLHYOGCOVQGNL 60
Db      1 MELAALCRMGILLALLPFGAASVQVCTGTMKRLPASPETHLDMRLHYOGCOVQGNL 60
Qy      61 ELTYLPTNASLSFLQDIQEVQGYVLIHNVQVQLQRLRIVRGTQLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASLSFLQDIQEVQGYVLIHNVQVQLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy      121 DPLNNTPTVAGASPGGIRELQLSLTETILKGVYLQNPOLCYQDTLLMWDIFRKNQOLA 180
Db      121 DPLNNTPTVAGASPGGIRELQLSLTETILKGVYLQNPOLCYQDTLLMWDIFRKNQOLA 180
Qy      181 LTLIDNRSRACHPCSPCKSGRSCWGSSESDCSLRTVAGAGCARCKGLPTDCCHEQC 240
Db      181 LTLIDNRSRACHPCSPCKSGRSCWGSSESDCSLRTVAGAGCARCKGLPTDCCHEQC 240
Qy      241 AAGCTGKMSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db      241 AAGCTGKMSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy      301 YNYLSTVGSCTLVCPILHNOETVABDQTCCKSCPKARVCYGLGMEHREYRAVTSAN 360
Db      301 YNYLSTVGSCTLVCPILHNOETVABDQTCCKSCPKARVCYGLGMEHREYRAVTSAN 360
Qy      361 ELEPAGCKIFGSLAPLPESFGDDPASNTAPLOPEQLQVETLEBITGYLISAMPDULP 420

```



```

Db      361 IOFACCKIFGSLAFLEPSFDGSDPASNTAPLQPEOLQVFEETLEITGYLYISAMPDLSL 420
Qy      421 DLSVFONLQVIRGRILHNGAYSITLQGLISWLGRLSELGSGALIHNNHLCFVHTV 480
Db      421 DLSVFONLQVIRGRILHNGAYSITLQGLISWLGRLSELGSGALIHNNHLCFVHTV 480
Qy      481 PMDQLFRNPHQALHTANREDECEVGEGLACHOLCRGHWGMPQCVCNCSQFLRGQEC 540
Db      481 PMDQLFRNPHQALHTANREDECEVGEGLACHOLCRGHWGMPQCVCNCSQFLRGQEC 540
Qy      541 VEECRVLOGLPREYVNAHCLPCHPECPQNSVTCFGEADQVCAHAHYKDPFCVAC 600
Db      541 VEECRVLOGLPREYVNAHCLPCHPECPQNSVTCFGEADQVCAHAHYKDPFCVAC 600
Qy      601 PFNNFTVSEFLARPKVASHL-----PDEGACOPICHTSCVDLDDKGCAPABRASP 656
Db      601 PFNNFTVSEFLARPKVASHL-----PDEGACOPICHTSCVDLDDKGCAPABRASP 656
Qy      657 LRSIVSAVGLLVVVGIVVFGILIKRQOKIRKYTRRLQETELVEBLTSGAMPNQA 716
Db      651 LRSIVSAVGLLVVVGIVVFGILIKRQOKIRKYTRRLQETELVEBLTSGAMPNQA 710
Qy      717 QMRILKETELRKXUGSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANEILDE 776
Db      711 QMRILKETELRKXUGSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANEILDE 770
Qy      777 AAYMAGVSPYVSRLLGICLTSTVQLVQLMPYGLLDHRENRGLSGODLLNMCQJA 836
Db      771 AAYMAGVSPYVSRLLGICLTSTVQLVQLMPYGLLDHRENRGLSGODLLNMCQJA 830
Qy      837 KGSYLEDVRLVRLDLAARNVLYKSPNHKITYDFGIALIDDETYHADGKVPKXMA 896
Db      831 KGSYLEDVRLVRLDLAARNVLYKSPNHKITYDFGIALIDDETYHADGKVPKXMA 890
Qy      897 LESILRRFTHQSDVMSYGVTVMLTFFGAKPYDGI PAEIPDLLEKGERLPOPICTID 956
Db      891 LESILRRFTHQSDVMSYGVTVMLTFFGAKPYDGI PAEIPDLLEKGERLPOPICTID 950
Qy      957 VYIMWKCMWIDSECRPRELIVSEFSRMAADPQRFVYIQNEGLGASPLDSTFYSLLE 1016
Db      951 VYIMWKCMWIDSECRPRELIVSEFSRMAADPQRFVYIQNEGLGASPLDSTFYSLLE 1010
Qy      1017 DDMGDLVAEEELVQCGFCPCDPAFGAGMWHHRSSSTPSGGGDLTLGIEPEEEA 1076
Db      1011 DDMGDLVAEEELVQCGFCPCDPAFGAGMWHHRSSSTPSGGGDLTLGIEPEEEA 1070
Qy      1077 PRSPLAPSEGAGSDVFDGLGMAKGLQSLPTHDESPLOQYSEDPVLPSETDGYVAP 1136
Db      1071 PRSPLAPSEGAGSDVFDGLGMAKGLQSLPTHDESPLOQYSEDPVLPSETDGYVAP 1130
Qy      1137 LTCSFQPEYVNOVDVPPQSPREGLPAPAPAGATLERAKTISFGKGVYKDVAFGGA 1196
Db      1131 LTCSFQPEYVNOVDVPPQSPREGLPAPAPAGATLERAKTISFGKGVYKDVAFGGA 1190
Qy      1197 VENPEYLTPOGAAPQHPAPSPAFDNLVYWDQPPERGAAPSTFKGTPAENEYLG 1256
Db      1191 VENPEYLTPOGAAPQHPAPSPAFDNLVYWDQPPERGAAPSTFKGTPAENEYLG 1250
Qy      1257 LDVPV 1261
Db      1251 LDVPV 1255

```

RESULT 13

AA051143
ID AA051143 standard; Protein; 1255 AA.

AC AA051143;

DT 17-JUN-2002 (first entry)

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

```

KW      Her-2/neu: oncogene; cancer; tumour; vaccine; human; p185;
KM      tyrosine kinase; receptor; c-erbB2; gene therapy.
XX      Homo sapiens.
OS      Homo sapiens.
FH      Key
FT      Location/Qualifiers
FT      1..653
FT      /note="extracellular domain"
FT      676..1255
FT      /note="intracellular domain"
FT      990..1255
FT      /note="phosphorylation domain"
PN      Domain
PD      WO200212341-A2.
PE      14-FEB-2002.
PF      03-AUG-2001; 2001WO-US24283.
PI      03-AUG-2000; 2000US-0632507.
PA      (CORI-) CORIYA CORP.
PA      (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
PI      Cheever MA, Gheysen D;
PI      WPI, 2002-241743/29.
PI      N-PESDB; ABA92250.
PT      Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT      or enhancing an immune response to the protein, has Her-2/neu
PT      extracellular domain fused to Her-2/neu intracellular or
PT      phosphorylation domain
PS      Claim 68; Fig 7; 141pp; English.
XX      The present sequence is that of human Her-2/neu (p185 glycoprotein
XX      or c-erbB2), an oncogenic self-protein and target for anti-cancer
XX      vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
XX      in a variety of cancers, including breast, ovarian, colon, lung and
XX      prostate cancer. Her-2/neu is a member of the tyrosine kinase
XX      family of receptor-like glycoproteins. It comprises an extracellular
XX      domain with homology to the epidermal growth factor receptor
XX      (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
XX      intracellular domain that also shows homology to EGFR. Its
XX      overexpression correlates with a poor prognosis in breast and
XX      ovarian cancers. The invention provides Her-2/neu fusion
XX      proteins, nucleic acids encoding them, viral vectors, and vaccines
XX      comprising the fusion proteins or nucleic acid molecules. In
XX      preferred fusion proteins, the extracellular domain of a Her-2/neu
XX      protein is fused to a Her-2/neu intracellular domain or
XX      phosphorylation domain (or its DeltaCD fragment). An immune
XX      response to Her-2/neu protein is elicited or enhanced by
XX      administering the fusion protein in the form of a vaccine, or by
XX      transfecting cells of an animal ex vivo with a nucleic acid
XX      encoding the fusion protein, and delivering the transfected cells
XX      to the animal. The fusion proteins, nucleic acids, and isolated
XX      specific T-cells are useful for inhibiting the development of a
XX      cancer, especially breast, ovarian, colon, lung or prostate cancer
XX      in a patient. T cells that specifically react with a Her-2/neu
XX      fusion protein can be used to remove tumour cells from a sample in
XX      order to inhibit the development of cancer in a patient.
SQ      Sequence 1255 AA;

```

Query Match

Best Local Similarity 96.8%; Score 6621; DB 23; Length 1255;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

Qy 1 MELALACRWGILLALLPPGASTGVCTGTDMKRLPASPETHLDMKRLHYGCGVVGNTL 60
Db 1 MELALACRWGILLALLPPGASTGVCTGTDMKRLPASPETHLDMKRLHYGCGVVGNTL 60

```

QY 61 ELTYLPTNASLSLQDIOEVQGYVLLIAHNVQROVPLQRLIRIVGTQLFEDNVALAVLDNG 120
DB 61 ELTYLPTNASLSLQDIOEVQGYVLLIAHNVQROVPLQRLIRIVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGRLQRLSTLEIIKGGVLIQRNPOLCYQDITLWKDIFHKNOQA 180
DB 121 DPLNNTTPVTGASPGGRLQRLSTLEIIKGGVLIQRNPOLCYQDITLWKDIFHKNOQA 180
QY 181 LTLIDNRSRACHPCSPCKGSRCWGESSEDCOSLTRVCAGGACAKGRLPTDCHEQC 240
DB 181 LTLIDNRSRACHPCSPCKGSRCWGESSEDCOSLTRVCAGGACAKGRLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESMNPBGRYTFGASCTYACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESMNPBGRYTFGASCTYACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVAEDTQCEKSPCARVCGLGMOYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPHNOEVAEDTQCEKSPCARVCGLGMOYIKANSKFIGIT 360
QY 361 ELFFAGCKITFGSLAFLPESFGDDPASNTAPLOPEOLQVETLEEITGYLYTSAMPDSL 420
DB 361 IOEFAGCKITFGSLAFLPESFGDDPASNTAPLOPEOLQVETLEEITGYLYTSAMPDSL 420
QY 421 DLSVFONLOVIRGRILHNGAYSLTLQGLISWLGASLEIGSGALIHNTHLCTVHTV 480
DB 421 DLSVFONLOVIRGRILHNGAYSLTLQGLISWLGASLEIGSGALIHNTHLCTVHTV 480
QY 481 PMDOLFRNPHOALHTANREDECEVGEGLACHOLCARGHGWGPTQVCNCSQFLRGQEC 540
DB 481 PMDOLFRNPHOALHTANREDECEVGEGLACHOLCARGHGWGPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAARHCLPCHPECOQNGSVTFGEBAOCVACAHYKXDPFVAVAC 600
DB 541 VEECRVLOGLPREYVNAARHCLPCHPECOQNGSVTFGEBAOCVACAHYKXDPFVAVAC 600
QY 601 PENNFVSFWLRYPKVYASHL-----PDEGACQCPINCTHSQVLDLKGCPAEGRASP 656
DB 601 PEG-----VKRDSYMPWKPEDEGACQCPINCTHSQVLDLKGCPAEGRASP 650
QY 657 LSIYSAVVGLLVVAVFGLILIRROQKIRKYMRLLOETLVEBLPSPGAMPNQA 716
DB 651 LSIISAVVGLLVVAVFGLILIRROQKIRKYMRLLOETLVEBLPSPGAMPNQA 710
QY 717 QMRILKETELRKVYVLSGAFGVYVGIWIPGENVKIPVALIKULRENTSPANKETIDE 776
DB 711 QMRILKETELRKVYVLSGAFGVYVGIWIPGENVKIPVALIKULRENTSPANKETIDE 770
QY 777 AYVMAVGSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSODLLANCMQIA 836
DB 771 AYVMAVGSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSODLLANCMQIA 830
QY 837 KGMSTYEDRLVHRDLAARNVLKSPNHKIDTDFGLARLLDDEIHYHADGKRVIKMA 896
DB 831 KGMSTYEDRLVHRDLAARNVLKSPNHKIDTDFGLARLLDDEIHYHADGKRVIKMA 890
QY 897 LESILRRRTTHQSDVWSYGVTWELMTFGAKPYDGPAREIPDLLEKGRPLPQPICTID 956
DB 891 LESILRRRTTHQSDVWSYGVTWELMTFGAKPYDGPAREIPDLLEKGRPLPQPICTID 950
QY 957 VYMWIKCMYIDSEGRPRFRELVSERSMARDPQRFVITQNEEDLCPASPLDSTFYRSILE 1016
DB 951 VYMWIKCMYIDSEGRPRFRELVSERSMARDPQRFVITQNEEDLCPASPLDSTFYRSILE 1010
QY 1017 DDDMGDLVDABEYLYPQOGFPDPAAPGAGVHRHRSSTRSGGSLTLGLEPSEBEA 1076
DB 1011 DDDMGDLVDABEYLYPQOGFPDPAAPGAGVHRHRSSTRSGGSLTLGLEPSEBEA 1070
QY 1077 PRSPLAPSEAGSDVFDGDLGGAAGKGLQSLPTHPSPLOYSSEDPVPLPSETGYVAP 1136
DB 1071 PRSPLAPSEAGSDVFDGDLGGAAGKGLQSLPTHPSPLOYSSEDPVPLPSETGYVAP 1130
QY 1137 LTCSPQPEVYNQPDVRPQPPSPRBPPLPAARPAKTLERAKTILSGKXGVVXDVAFAGGA 1196

```

```

DB 1131 LTCSPQPEVYNQPDVRPQPPSPRBPPLPAARPAKTLERAKTILSGKXGVVXDVAFAGGA 1190
QY 1197 VENEPEYLTPOGGAAPHPHPAFSAFNLTYWODDPERKGAPESTFNGTPTAENPEYLIG 1256
DB 1191 VENEPEYLTPOGGAAPHPHPAFSAFNLTYWODDPERKGAPESTFNGTPTAENPEYLIG 1250
QY 1257 LDVFPV 1261
DB 1251 LDVFPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein, 1255 AA.
XX
XX AAU77114;
AC
XX 05-JUN-2002 (first entry)
DT
XX Human Her-2/neu polypeptide.
DE
XX
XX Human; Her-2/neu; cytostratic; haematological malignancy; CML;
XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
XX chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
XX Hodgkin's lymphoma; T cell therapy.
XX
XX Homo sapiens.
XX WO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX MPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
XX encoding the polypeptide, or antigen presenting cells expressing the
XX polypeptide.
XX
XX Disclosure; Page 71-74; 74pp; English.
XX
XX The invention relates to a method for inhibiting development of
XX haematological malignancy in a patient by administering a polypeptide
XX comprising an immunogenic portion of Her-2/neu or a polynucleotide
XX encoding the polypeptide. Antigen presenting cells that express the
XX protein can also be administered. The sequences are used for inhibiting
XX development of haematological malignancy such as acute myelogenous
XX leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
XX leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
XX lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
XX Sequence 1255 AA;
XX
XX Query Match 96.8%; Score 6621; DB 23; Length 1255;
XX Best Local Similarity 96.8%; Pred. No. 0;
XX Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
QY 1 MELAAICRMGLTALALPBGASSTQVCTGTMKRLPASPEPTHLMRLHYGGQGVYQGNL 60
DB 1 MELAAICRMGLTALALPBGASSTQVCTGTMKRLPASPEPTHLMRLHYGGQGVYQGNL 60
QY 61 ELTYLPTNASLSLQDIOEVQGYVLLIAHNVQROVPLQRLIRIVGTQLFEDNVALAVLDNG 120

```

```

Db      61  ELTYLPTNASTSLFLOIOEVQGYVLI.AHNQVRQVPLQRIYRGITQLFEDNYALALVDNG 120
Qy      121  DPLNNTTPTVTGASPGGARELOLSTLTKGAVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121  DPLNNTTPTVTGASPGGARELOLSTLTKGAVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181  LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGGACARCKGPLPTDCHEQC 240
Db      181  LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGGACARCKGPLPTDCHEQC 240
Qy      241  AAGCTGPKHSDCLACI.HFNHSGI.CELHCPALVTYNTDTESMNPREGRTYFGASCYTACP 300
Db      241  AAGCTGPKHSDCLACI.HFNHSGI.CELHCPALVTYNTDTESMNPREGRTYFGASCYTACP 300
Qy      301  YNYLSTDVGSCTLVCPPLNQEVTAEQDTCRCKSPCARVCYGLMGOYIKANSKEFIGIT 360
Db      301  YNYLSTDVGSCTLVCPPLNQEVTAEQDTCRCKSPCARVCYGLMGOYIKANSKEFIGIT 360
Qy      361  ELFEAGCKTI.FGSLAF.PESFDDPASNTPALQPELOVFEETLEETGYLYISAMPDISP 420
Db      361  ELFEAGCKTI.FGSLAF.PESFDDPASNTPALQPELOVFEETLEETGYLYISAMPDISP 420
Qy      421  DLSVFONTQYIRGRILHNGAYSLTLOGLISWLGRLSRLSGSLAL.IHNNHLCFVHTV 480
Db      421  DLSVFONTQYIRGRILHNGAYSLTLOGLISWLGRLSRLSGSLAL.IHNNHLCFVHTV 480
Qy      481  PMDOLFRNPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPQOCVNCQOPLRGQEC 540
Db      481  PMDOLFRNPHQALLHTANRPEDCEVGEGLACHQLCARGHCWGPQOCVNCQOPLRGQEC 540
Qy      541  VEECRVLOGLPREYVNAARHCLPCHPEQOPNGSVTCFGEADQCVACAHYKDPFCEVAC 600
Db      541  VEECRVLOGLPREYVNAARHCLPCHPEQOPNGSVTCFGEADQCVACAHYKDPFCEVAC 600
Qy      601  PENNFVSWLRVPCVSAHLE---PDEGACQPCPINCTHSCVDLDOKCPAEGRASBP 656
Db      601  PSC-----VKSDLSYMPIMKPFDEGACQPCPINCTHSCVDLDOKCPAEGRASBP 650
Qy      657  LTSIVSAVVGILLVWVGVFGILIRKROOKIRKYMRRLOJTELEVEPLTPSGAMPNA 716
Db      651  LTSIISAVGILLVWVGVFGILIRKROOKIRKYMRRLOJTELEVEPLTPSGAMPNA 710
Qy      717  QMRILKETEELRKVJGSGAFGVYKGIWIPDENYKIPVAIYKLEBNTSPKANKELIDE 776
Db      711  QMRILKETEELRKVJGSGAFGVYKGIWIPDENYKIPVAIYKLEBNTSPKANKELIDE 770
Qy      777  AAYMAGVGSFYVSRLLGICLTSTVOLVTOLMPYGCILLDHRENRGLSGODLNNCMQIA 836
Db      771  AAYMAGVGSFYVSRLLGICLTSTVOLVTOLMPYGCILLDHRENRGLSGODLNNCMQIA 830
Qy      837  KGSVYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDIDETEYHADGGKVPYIKMA 896
Db      831  KGSVYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDIDETEYHADGGKVPYIKMA 890
Qy      897  LESILRRRFTHQSDWSYGVYTWELMTFGAKPYDGIIPAREIPLLLEKGERLPPPTCTID 956
Db      891  LESILRRRFTHQSDWSYGVYTWELMTFGAKPYDGIIPAREIPLLLEKGERLPPPTCTID 950
Qy      957  VYMIIMYKMMIDSECRPFRELVESEFSMARDPQRFVJQNEDELGASPLDSTFYSLLE 1016
Db      951  VYMIIMYKMMIDSECRPFRELVESEFSMARDPQRFVJQNEDELGASPLDSTFYSLLE 1010
Qy      1017  DDMGDLVDAEYIYVPOGFCFCDPAPAGAGMYHHNRSSSTSGGDDLTLGLEPSEEEA 1076
Db      1011  DDMGDLVDAEYIYVPOGFCFCDPAPAGAGMYHHNRSSSTSGGDDLTLGLEPSEEEA 1070
Qy      1077  PRSPLASEGAGSVDFGDLGMMGAKGLOSLPTHDSPLQRYSEDETPVLPSESTDGYVAP 1136
Db      1071  PRSPLASEGAGSVDFGDLGMMGAKGLOSLPTHDSPLQRYSEDETPVLPSESTDGYVAP 1130
Qy      1137  LTGSPQPEYVNOQDVRPQPSPREGPIPAARPAATLIERAKTLSPKNGVYKQVAFRGA 1196

```

```

Db      1131  LTGSPQPEYVNOQDVRPQPSPREGPIPAARPAATLIERAKTLSPKNGVYKQVAFRGA 1190
Qy      1197  VENEPEYLTPOGGAAPQPHPPAFSPAFDNLYWMDPPERGAPOSTKGTFTANPEYLG 1256
Db      1191  VENEPEYLTPOGGAAPQPHPPAFSPAFDNLYWMDPPERGAPOSTKGTFTANPEYLG 1250
Qy      1257  LDVPEY 1261
Db      1251  LDVPEY 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein: 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KM Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN MO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA) CREATIVE BIOMOLECULES INC.
PI Houston TX, Huston JS, Oppermann H, Ring DB,
XX
XX WPI: 1993-272889/34.
XX
XX N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 96.2%; Score 6578; DB 14; Length 1433;
Best Local Similarity 96.3%; Freq. No. 0;
Matches 1218; Conservative 9; Mismatches 24; Indels 14; Gaps 2;

Qy      1  MELALCRMGILLALPPGAASVCTGTDMLRLPASPETHLMRLLYOGCVQVQNTL 60
Db      1  MELALCRMGILLALPPGAASVCTGTDMLRLPASPETHLMRLLYOGCVQVQNTL 60
Qy      61  ELTYLPTNASTSLFLOIOEVQGYVLI.AHNQVRQVPLQRIYRGITQLFEDNYALALVDNG 120
Db      61  ELTYLPTNASTSLFLOIOEVQGYVLI.AHNQVRQVPLQRIYRGITQLFEDNYALALVDNG 120
Qy      121  DPLNNTTPTVTGASPGGARELOLSTLTKGAVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db      121  DPLNNTTPTVTGASPGGARELOLSTLTKGAVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy      181  LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGGACARCKGPLPTDCHEQC 240
Db      181  LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGGACARCKGPLPTDCHEQC 240

```

QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTEESMPNEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTEESMPNEGRYTFGASCVTACP 300
QY 301 VNYLSTDVSGCTLVCPHNOEYTAEDGTORCEKSKPCARVCYGLQMOYIKANSKFIGIT 360
DB 301 VNYLSTDVSGCTLVCPHNOEYTAEDGTORCEKSKPCARVCYGLQMEHLREVRAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVETLEETGTYLISAMPDSLP 420
DB 361 IGEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEHLQVETLEETGTYLISAMPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSLELGSGLALIHNTHLCPVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSLELGSGLALIHNTHLSPVHTV 480
QY 481 PMDQLEFRNPHQALHTANRPEDECEVCEGLACHQOLCARGHCMGPPTQCVCNCSQFLRQEC 540
DB 481 PMDQLEFRNPHQALHTANRPEDECEVCEGLACHQOLCARGHCMGPPTQCVCNCSQFLRQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPQHPECQPNQSVTCFGEADQVCAHAYKDPPECVAC 600
DB 541 VEECRVLOGLPREYVNAHCLPQHPECQPNQSVTCFGEADQVCAHAYKDPPECVAC 600
QY 601 PNNFTVSFWLVRPKVASHLE---PDEEGACOPCPINCHSCVDLDDKCGPAEQRAS 656
DB 601 PGG-----VKPDLSTYMPIMKFPDEEGACOPCPINCHSCVDLDDKCGPAEQRAS 650
QY 657 LTSIVSAVVGLLVVYLVGVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOA 716
DB 657 LTSIVSAVVGLLVVYLVGVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDE 776
DB 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDE 770
QY 777 AYVAGVSGPYVSRLLIGICTSTVOLVTLQMPYGLDHYRENRGLSODLNMCMOIA 836
DB 777 AYVAGVSGPYVSRLLIGICTSTVOLVTLQMPYGLDHYRENRGLSODLNMCMOIA 830
QY 837 KGMSTYLEDVRLVHRLDLAARNVLVKS PNHVKITDFGLARLIDETEHADGKVPKXMA 896
DB 837 KGMSTYLEDVRLVHRLDLAARNVLVKS PNHVKITDFGLARLIDETEHADGKVPKXMA 890
QY 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PARETPDLLEKGERLPQPTCTID 956
DB 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PARETPDLLEKGERLPQPTCTID 950
QY 957 VYMIWKCMIIDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLE 1016
DB 957 VYMIWKCMIIDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLE 1010
QY 1017 DDDMGDDVDAERYLYPQGGFFCPDPAFGAGVTHHRSSSTRSGGDLTLGLEPSEEEA 1076
DB 1017 DDDMGDDVDAERYLYPQGGFFCPDPAFGAGVTHHRSSSTRSGGDLTLGLEPSEEEA 1070
QY 1077 PRSPPLAPSEAGSDVFPDGLGGAAGLQSLPTHDPSPLOXYSEDPVPLPSPTDGYVAP 1136
DB 1077 PRSPPLAPSEAGSDVFPDGLGGAAGLQSLPTHDPSPLOXYSEDPVPLPSPTDGYVAP 1130
QY 1137 LTCSPQPEYVNOQDVRFQPPSPREGPLPAAPAGATLERAKTLSPGXGVYKDVAFAGGA 1196
DB 1137 LTCSPQPEYVNOQDVRFQPPSPREGPLPAAPAGATLERAKTLSPGXGVYKDVAFAGGA 1190
QY 1197 VENPEYLPQGGAAQPHPPAFSPAFNDLYWMDODPERGAPSTFKGTATENPEYLG 1256
DB 1197 VENPEYLPQGGAAQPHPPAFSPAFNDLYWMDODPERGAPSTFKGTATENPEYLG 1250
QY 1257 LDVPV 1261
DB 1257 LDVPV 1255

Search completed: July 22, 2003, 08:41:51
Job time : 44.1589 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds

(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-632-652-14

Perfect score: 6834

Sequence: 1 MELALCRMGALLALLPPCA.....TRKGPPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database:

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6623	96.9	1255	1 A24571	protein-tyrosine k
2	5864	85.8	1260	1 TVRNU	protein-tyrosine k
3	5841.5	85.5	1254	1 I48161	p-185 precursor
4	3113	45.6	1210	1 GQHDE	epidermal growth f
5	3080	45.1	1210	1 A53183	epidermal growth f
6	3049.5	44.6	1223	1 TVCHLV	epidermal growth f
7	2920.5	42.7	1308	1 A47253	epidermal growth f
8	2627	38.4	1166	1 S06142	protein-tyrosine k
9	2389.5	35.0	1342	2 A36223	kinase-related tra
10	2302.5	33.7	1339	2 UC487	epidermal growth f
11	1721.5	25.2	698	1 TVFVLV	epidermal growth f
12	1658	24.3	604	1 TVYOH	protein-tyrosine k
13	1611.5	23.6	1330	1 GQFPE	epidermal growth f
14	1602	23.4	544	2 S35745	protein-tyrosine k
15	1595	23.3	545	2 S00727	kinase-related tra
16	1578	23.1	540	2 B44776	protein-tyrosine k
17	1576	23.1	540	2 TVFVEB	epidermal growth f
18	1509	22.1	644	2 A36325	epidermal growth f
19	1274	18.6	1333	2 E88257	protein-tyrosine k
20	1274	18.6	1374	2 S70712	protein-tyrosine k
21	1195	17.5	1369	2 S70713	protein-tyrosine k
22	1150	16.8	1717	1 A45558	epidermal growth f
23	1136	16.5	527	2 A42032	epidermal growth f
24	975.5	14.3	893	2 A27331	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	713	10.4	1363	2 T43220	insulin-like growt
28	694	10.2	1382	1 INHUR	insulin receptor p
29	693.5	10.1	1372	2 A34157	insulin receptor p

30	690	10.1	1383	2 A36080	insulin receptor p
31	680	10.0	1477	2 T18534	protein-tyrosine k
32	680	10.0	1607	2 T43212	insulin-like growth
33	679.5	9.9	1300	2 A36502	insulin receptor-r
34	669	9.8	1268	2 B36502	insulin receptor-r
35	644	9.4	1367	1 IGHUR1	insulin-like growth
36	629	9.2	1371	2 A33837	insulin-like growth
37	622.5	9.1	2148	1 A56081	insulin receptor -
38	617	9.0	1390	2 T30346	insulin receptor -
39	616	9.0	2101	2 S57245	insulin receptor (
40	591	8.6	1114	1 S05582	protein-tyrosine k
41	589.5	8.6	1028	2 A45388	protein-tyrosine k
42	586	8.6	1091	2 S33596	protein-tyrosine k
43	579	8.5	1052	2 A46166	protein-tyrosine k
44	578.5	8.5	952	2 I50612	protein-tyrosine k
45	576.5	8.4	984	2 A39753	protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N.Aleternate names: c-erbB-2 protein precursor; kinase-related transforming protein C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erbB-2 gene to epidermal gro

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/ep

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seebu

Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares ch

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

A/Residue: 1-517, 'RALV', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Residue: 1-517, 'RALV', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; MUID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcript:

A/Reference number: I57622; MUID:87286898; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:g181983; PIDN:AA5637.1; PID:g553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetics:
 A:Gene: GDB:ERBB2; NGI; NEU; HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the list of introns is incomplete
 C:Function:
 A:Description: catalyzes the phosphorylation of a peptide tyrosine residue by ATP
 A:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:22-653/Domain: extracellular #status predicted <EXT>
 F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>
 F:395-605/Domain: EGF receptor extracellular domain repeat <EB2>
 F:654-675/Domain: transmembrane #status predicted <TM>
 F:676-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Region: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:68,124,187,259,530,571,629/Binding site: carboxylate (asn) (covalent) #status predicted
 F:686/Binding site: phosphate (thr) (covalent) (by protein kinase C) #status predicted
 F:753/Active site: Lys #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.9% Score 6623; DB 1; Length 1255;
 Best local similarity 97.3% Pred. No. 3.6e-255;
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALPFGAASVCTGTDMKRLPASPETHLDMLEHLYOGCCVQVQNT 60
 DB 1 MELAALCRWGLLLALPFGAASVCTGTDMKRLPASPETHLDMLEHLYOGCCVQVQNT 60
 QY 61 ELTLPNTMASLSFQDIOENOGVYLIAHQVRLQRLRYRGTQCFEDNYALAVINDG 120
 DB 61 ELTLPNTMASLSFQDIOENOGVYLIAHQVRLQRLRYRGTQCFEDNYALAVINDG 120
 QY 121 DPLNNTPTVGTASPGGLRELQSLTEILKGVLIQENPOLCYODTILMKDIFKNNQLA 180
 DB 121 DPLNNTPTVGTASPGGLRELQSLTEILKGVLIQENPOLCYODTILMKDIFKNNQLA 180
 QY 181 LTLIDTNRSPACHPCSPCKSKRCWGESSEDDCSITRTVCAGGACRKGPIPTDCHEQC 240
 DB 181 LTLIDTNRSPACHPCSPCKSKRCWGESSEDDCSITRTVCAGGACRKGPIPTDCHEQC 240
 QY 241 AAGCTGPRGSDCLACLNHSGICELHCPALVTYNTDTFESMPREGRYTFGASCTYACP 300
 DB 241 AAGCTGPRGSDCLACLNHSGICELHCPALVTYNTDTFESMPREGRYTFGASCTYACP 300
 QY 301 YNYLSTDVGSCTTVCPILHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360
 DB 301 YNYLSTDVGSCTTVCPILHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360
 QY 301 YNYLSTDVGSCTTVCPILHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360
 DB 301 YNYLSTDVGSCTTVCPILHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360
 QY 361 ELLEFAGCKKIGSLAFIPESFDGPASNTAPLOEQLOVEFTLEITGYLYISWPSLP 420
 DB 361 ELLEFAGCKKIGSLAFIPESFDGPASNTAPLOEQLOVEFTLEITGYLYISWPSLP 420
 QY 421 DLSVFNQLOVIRGRILHNGAYSLTQGLIGISWLGRLSRELGSGLALIHNTTHLCFVHTV 480
 DB 421 DLSVFNQLOVIRGRILHNGAYSLTQGLIGISWLGRLSRELGSGLALIHNTTHLCFVHTV 480
 QY 481 PMDQLFRRPHQALLHTANRPEDECVGEGELACHQICARHCHGPRPTQCVNSQFLRQEC 540
 DB 481 PMDQLFRRPHQALLHTANRPEDECVGEGELACHQICARHCHGPRPTQCVNSQFLRQEC 540
 QY 541 VEEGRVLOGLPREVYNAHCLPCHPECCPQNGSVTCFEPADQCVACAHYDPFCVARC 600
 DB 541 VEEGRVLOGLPREVYNAHCLPCHPECCPQNGSVTCFEPADQCVACAHYDPFCVARC 600
 QY 601 PSQVPRDLSTYMPKPEDEGACQPCPNCTHSCVDLDDKCCPAPORASPLTISFNFTV 660
 DB 601 PSQVPRDLSTYMPKPEDEGACQPCPNCTHSCVDLDDKCCPAPORASPLTISFNFTV 660

DB 601 PSQVPRDLSTYMPKPEDEGACQPCPNCTHSCVDLDDKCCPAPORASPLTISAVVG 660
 QY 661 SEFLRVPKYSASHLEKRRQCKTRKTYMRLLLOETELVELTSSGAMPNDAQRIKLTETL 720
 DB 661 SEFLRVPKYSASHLEKRRQCKTRKTYMRLLLOETELVELTSSGAMPNDAQRIKLTETL 720
 QY 721 RKVKVLGSAFGVVKGIWI PDGENVKIYVAIKVLRNTPANKKEILDEAVYAGVSGP 780
 DB 721 RKVKVLGSAFGVVKGIWI PDGENVKIYVAIKVLRNTPANKKEILDEAVYAGVSGP 780
 QY 781 YVSRLLGICLTSTVQLVTLQMPYGCILDHVRENRGLSGQDILNMCQIAKMSYLEVR 840
 DB 781 YVSRLLGICLTSTVQLVTLQMPYGCILDHVRENRGLSGQDILNMCQIAKMSYLEVR 840
 QY 841 LVHRDLAARNVYKSNHKKINDFGIARLLIDETEVADGGKVPKMMALSLIRRRFT 900
 DB 841 LVHRDLAARNVYKSNHKKINDFGIARLLIDETEVADGGKVPKMMALSLIRRRFT 900
 QY 901 HOSDWSYGVTVWELMTFGAKPYDGI PAREIDLLKGERLPOPPICITIDVYIMVKCM 960
 DB 901 HOSDWSYGVTVWELMTFGAKPYDGI PAREIDLLKGERLPOPPICITIDVYIMVKCM 960
 QY 961 IDSECRPRRELVSERSMARDPQRFVYQNEIDGPAFLDSTFRRSLLEDMDGLVVA 1020
 DB 961 IDSECRPRRELVSERSMARDPQRFVYQNEIDGPAFLDSTFRRSLLEDMDGLVVA 1020
 QY 1021 EETLVPOQGFPCDPAPAGWYHHRSSSTRSGGQDITLGLPESEAPRSPPLAPSEG 1080
 DB 1021 EETLVPOQGFPCDPAPAGWYHHRSSSTRSGGQDITLGLPESEAPRSPPLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPQEVY 1140
 DB 1081 AGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPQEVY 1140
 QY 1141 NOPDVRPOPSPREGLPAPAPGATLEAKTSLPCKNVWDVFAFGAVNPEVLTQ 1200
 DB 1141 NOPDVRPOPSPREGLPAPAPGATLEAKTSLPCKNVWDVFAFGAVNPEVLTQ 1200
 QY 1201 GGAAPQHPHPAPAFDNLVYWDQPPERGAPEPTFKTPTAENPEVGLDVPV 1255
 DB 1201 GGAAPQHPHPAPAFDNLVYWDQPPERGAPEPTFKTPTAENPEVGLDVPV 1255

RESULT 2
 TVRTNU
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related prote
 A:Reference number: A24562; MUID:86118662; PMID:3945311
 A:Accession: A24562
 A:Molecule type: mRNA
 A:Residues: 11260 <BAR>
 A:Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
 R:Masu, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cc
 Carcinogenesis 12, 1975-1978, 1991
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals
 2-thiacyloyl]formamide or N-methyl-N-nitrosourea.
 A:Reference number: A61204; MUID:92035293; PMID:1682063
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663; 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-660/Domain: transmembrane #status predicted <TM>
 F:723-968/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:71-191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted
 F:758/Active site: lys #status predicted
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.8%; Score 5864; DB 1; Length 1260;
 Best Local Similarity 86.2%; Pred. No. 5,7e-234;
 Matches 1084; Conservative 48; Mismatches 123; Indels 2; Gaps 2;

QY 1 MELALCRWGLLALLPPGAASVCTGCTDMKRLPASPETHLDMLRLHYOGCVVQGNL 60
 DB 4 MELAMCRWGLLALLPPGIAGTQCTGDMKRLPASPETHLDMLRLHYOGCVVQGNL 63
 QY 61 ELTYLPTNASISFLDDIOEVQGYVLIANOVQVPLRLIRVGTQLFEDNYALAVLDNG 120
 DB 64 ELTYVPANASISFLDDIOEVQGYVLIANOVQVPLRLIRVGTQLFEDNYALAVLDNR 123
 QY 121 DPLNNTPTVT-GASPGRLREIQRLSLTEILKGGVLIQRNPQLCYODTILMKDI FHKNNQL 179
 DB 124 DPODNVAASPPGRTPBGLREIQRLSLTEILKGGVLIQRNPQLCYODVILMKDVFRRKNNQL 183
 QY 180 ALTLIDTNRSPACHPCSPMKGSRCWGESSEDCOSLTRVYACAGACRCKGLPTDCCHQ 239
 DB 184 APVIDITNRSPACHPCSPACKDNHGWGESSEDCOILITITCTGACRCKGLPTDCCHQ 243
 QY 240 CAAGCTGPKHSDCLACJHFNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTAC 299
 DB 244 CAAGCTGPKHSDCLACJHFNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTTC 303
 QY 300 PYNLTSDVGSCTLYVCLHNOEYTAEDGTORCEKSKPCARVCYGLGMOYITKANSFICG 359
 DB 304 PYNLTSTEVGSCTLYVCPNNOEYTAEDGTORCEKSKPCARVCYGLGMEHLRGARAITSD 363
 QY 360 TELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVETLEETITGLYISAMPDGL 419
 DB 364 NVOEFOSCKKIFGSLAPLPESFDGDPSSGIALPPELOVETLEETITGLYISAMPDGL 423
 QY 420 PDLSVFONLQVIRGRILHNGAYSILTLQGLISWLGRLSRELGSGALILHNTLHLCFVHT 479
 DB 424 RDLVSFONLRIIRGRILHNGAYSILTLQGLISWLGRLSRELGSGALILHNTLHLCFVHT 483
 QY 480 VPMDOLFERNPHOALLHTANRPEDE-CYBEGLAGHOICARHCHGCPGPTQCVNCSOTLRQ 538
 DB 484 VPMDOLFERNPHOALLHSGNRPEEDLCVSSGLVCNSLCAHCHGCPGPTQCVNCSOTLRQ 543
 QY 539 ECVBECRVLQGLPREYVNAHCLPCHBECQPNQSVYTCFEPADQCVACAHYKDPFCVYA 598
 DB 544 ECVBECRVWKGGLPREYVSDKRLPCHBECQPNQSVYTCFEPADQCVACAHYKDPFCVYA 603
 QY 599 RCSSGVKPDLSYMPKPFDEBAGACQPCPINCNSCYDLDKCCPAPBORAPLTIS-FNNF 658
 DB 604 RCSSGVKPDLSYMPKPFDEBAGACQPCPINCNSCYDLDKCCPAPBORAPLTIS-FNNF 663
 QY 659 TVSFMDLRVPKVASHLERKROCKIRKYMRLLOETELVEPLTPSGAMPNOQOMRLKET 718
 DB 664 EGYLLFLIVVVGILIKRRROKIRKYMRLLOETELVEPLTPSGAMPNOQOMRLKET 723
 QY 719 ELARKVULSGAGTYKGIWIPDGENVKI PVAIKYLRENTSPKANKIILDEAYVAVAGV 778
 DB 724 ELARKVULSGAGTYKGIWIPDGENVKI PVAIKYLRENTSPKANKIILDEAYVAVAGV 783
 QY 779 SPVYSRLGICLSTYQVLTQMLPYPGCLDHYRENGRGLSGODLWMCQIAKMSYLED 838
 DB 784 SPVYSRLGICLSTYQVLTQMLPYPGCLDHYRENGRGLSGODLWMCQIAKMSYLED 843
 QY 839 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPIMWALLESTLRR 898
 DB 844 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPIMWALLESTLRR 903
 QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLENGERLPPPTCTIDVYIMWKC 958

DB 904 FTHQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLENGERLPPPTCTIDVYIMWKC 963
 QY 959 WVIDSCRRPFEIUSFSEFRMRDQRFVYIQNEQLGSPALDSTFYRSLLEDDDDMDLV 1018
 DB 964 WVIDSCRRPFEIUSFSEFRMRDQRFVYIQNEQLGSPALDSTFYRSLLEDDDDMDLV 1023
 QY 1019 DAEEYLPQCGFFCCDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEPAPRSLAPS 1078
 DB 1024 DAEEYLPQCGFFSDPFTPGTSTAHRRSSSTRSGGDLTLGLEPSEEPAPRSLAPS 1083
 QY 1079 EGAGSDVPFGDILCMGAKLQSLPTHDSPLORYSEDPVLPSTETGVAPLTCSPQE 1138
 DB 1084 EGAGSDVPFGDILAMGVTKQLSLSPHDSPLORYSEDPVLPSTETGVAPLTCSPQE 1143
 QY 1139 YNNOPEPQPPSPREGLPAPRPAATLERAKTSLPGKNGVXDVAFAGAVENPEYLT 1198
 DB 1144 YNNOPEVQPPPLTEBGLPVPAPGATLERKXTSLPGKNGVXDVAFAGAVENPEYLT 1203
 QY 1199 PQGAAPQPPPPAPSPAFDNLVYWDODPPERGAPSPSTFKGTAEENPEYLGLDVPPV 1255
 DB 1204 PREGTASPPHSPAPSPAFDNLVYWDQNSSEGGPPSPNEGTFTAEENPEYLGLDVPPV 1260

RESULT 3

148161
 p-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C:Accession: 148161
 R:Nakamura, T.; Ushijima, T.; Iehizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishi
 Gene 140, 251-255, 1994
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Reference number: 148161; MID:94193007; PMID:7908275
 A:Accession: 148161
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1254 <RES>
 A:Cross-references: GB:D16295; NID:9493236; PID:BA003801.1; PID:9747595
 C:Genetics:
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.5%; Score 5841.5; DB 2; Length 1254;
 Best Local Similarity 85.7%; Pred. No. 4,7e-233;
 Matches 1076; Conservative 57; Mismatches 121; Indels 1; Gaps 1;

QY 1 MELALCRWGLLALLPPGAASVCTGCTDMKRLPASPETHLDMLRLHYOGCVVQGNL 60
 DB 1 MELAMCRWGLLALLPPGASGTQCTGDMKRLPASPETHLDMLRLHYOGCVVQGNL 60
 QY 61 ELTYLPTNASISFLDDIOEVQGYVLIANOVQVPLRLIRVGTQLFEDNYALAVLDNG 120
 DB 61 ELTYVPANASISFLDDIOEVQGYVLIANOVQVPLRLIRVGTQLFEDNYALAVLDNR 120
 QY 121 DPLNNTPTVTGASPGRLREIQRLSLTEILKGGVLIQRNPQLCYODTILMKDI FHKNNQL 180
 DB 121 DPLNNTPTVTGASPGRLREIQRLSLTEILKGGVLIQRNPQLCYODTILMKDI FHKNNQL 180
 QY 181 LTLIDTNRSPACHPCSPMKGSRCWGESSEDCOSLTRVYACAGACRCKGLPTDCCHQ 240
 DB 181 LTLIDTNRSPACHPCSPMKGSRCKHCGASPECCOTLTGTIAPRAVAPAAARLPTDCCHQ 240
 QY 241 AAGCTGPKHSDCLACJHFNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTAC 300
 DB 241 AAGCTGPKHSDCLACJHFNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTTC 300
 QY 301 PYNLTSDVGSCTLYVCLHNOEYTAEDGTORCEKSKPCARVCYGLGMOYITKANSFICG 360
 DB 301 PYNLTSTEVGSCTLYVCPNNOEYTAEDGTORCEKSKPCARVCYGLGMEHLRGARAITSAN 360

```

Qy 361 ELBAGCKKIFGSLAFIPESFDGDPASNTAPLOEOLVETTEITGYLYISAMPDLSL 420
Db 361 IQBAGCKKIFGSLAFIPESFDGDPASNTAPLOEOLVETTEITGYLYISAMPDLSL 420
Qy 421 DLSVFNOLVIRGRILHNAGAYSLTLOGISWLGRLSRLSGLALIHNTHLCEVHTV 480
Db 421 DLSVFNOLVIRGRILHNAGAYSLTLOGISWLGRLSRLSGLALIHNTHLCEVHTV 480
Qy 481 PMDQFRPHQALLHTANRPDECEVGEGLACHQCAHGCKGPERPTCCVNCOSQLRQEC 540
Db 481 PMDQFRPHQALLHTANRPDECEVGEGLACHQCAHGCKGPERPTCCVNCOSQLRQEC 540
Qy 541 VEGRVLIQGLPREVYNARHCLPCHPECOPQNGSVTCGPEADQCVACAHXDPFCVARC 600
Db 541 VKEGRVWKGLPREVYNGHCLPCHPECOPQNGSVTCGPEADQCVACAHXDPFCVARC 600
Qy 601 PSQVYKPLSTYMPIMKFPDEBGAQCPPIINCHSCVDDDKCCPAEQASPLTSTFNNFTV 660
Db 601 PSQVYKPLSTYMPIMKFPDEBGAQCPPIINCHSCVDDDKCCPAEQASPLTSTFNNFTV 660
Qy 661 SFMLRVKVSASHLEKSRQCKIRKYTRMRLLQETELVEPLTPSGAMNQOMRLKETEL 720
Db 661 ILLELVIGVAVGILIKRRQCKIRKYTRMRLLQETELVEPLTPSGAMNQOMRLKETEL 720
Qy 721 RKVYVLSGAFGTYVKGIMPDGNNVKIPVAIKYLRNTSPKANKELIDEAYVYVAGVGP 780
Db 721 RKVYVLSGAFGTYVKGIMPDGNNVKIPVAIKYLRNTSPKANKELIDEAYVYVAGVGP 780
Qy 781 YVSRLLGICLTSTVQVLTQMLPYGCLLDHVENENGRIGSOOLNMCQIKKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQVLTQMLPYGCLLDHVENENGRIGSOOLNMCQIKKMSYLEDVR 840
Qy 841 LVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEVHADGKVPIMKMALESILRRFT 900
Db 841 LVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEVHADGKVPIMKMALESILRRFT 900
Qy 901 HQSDVMSGYTVVLMFTFGAKFYDGIIPAREIPDLLENGERLPQPPICITIDVYMMVKCM 960
Db 901 HQSDVMSGYTVVLMFTFGAKFYDGIIPAREIPDLLENGERLPQPPICITIDVYMMVKCM 960
Qy 961 IDSECRPFRELVESEFSMARDPQRFVVIQNEIDGSPASPLDSTFYRSLDEDDMGDLVDA 1020
Db 961 IDSECRPFRELVESEFSMARDPQRFVVIQNEIDGSPASPLDSTFYRSLDEDDMGDLVDA 1020
Qy 1021 EBYLVPOQGFCCPPAPAGAGVMHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080
Db 1021 EBYLVPOQGFCCPPAPAGAGVMHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPTVPLPSETDGVVAPLTGSPQREYV 1140
Db 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPTVPLPSETDGVVAPLTGSPQREYV 1140
Qy 1141 NQDPVRPOPPEPRGGLPAPAPAGATLERAKTISPGKNGVYKDVAFAGAVENEPYLPQ 1200
Db 1141 NQDPVRPOPPEPRGGLPAPAPAGATLERAKTISPGKNGVYKDVAFAGAVENEPYLPQ 1200
Qy 1201 GGAAPOPHPAPFAPAFDNLVYMDQDPFERGAPSTFKGPTTANPEYLGDIVV 1255
Db 1201 GGAAPOPHPAPFAPAFDNLVYMDQDPFERGAPSTFKGPTTANPEYLGDIVV 1255

```

RESULT 4

GQNHU
 Epidermal growth factor receptor precursor - human
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbb
 C:Species: Homo sapiens (man)
 C:Date: 15-Nov-1994 #sequence_revision 27-Nov-1995 #text_change 11-Jun-1999
 C:Accession: A00641, A25772, A38672, A00642, A43615, A23062, A05281, A60143, A13
 R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y
 rg, P.H.
 A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312
 A:Molecule type: mRNA
 A:Residues: 1-1210
 A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA2540.1; PID:g757924
 A:Note: the authors translated the codon AAG for residue 540 as Aen
 R:Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
 Proc. Natl. Acad. Sci. U.S.A. 82: 4920-4924, 1985
 A:Title: Characterization and sequence of the promoter region of the human epiderma
 A:Reference number: A25772; MUID:85270438; PMID:2991899
 A:Accession: A25772
 A:Molecule type: DNA
 A:Status: translation not shown
 A:Residues: 1-29 <IGH>
 A:Cross-references: GB:M1234; NID:g181981; PIDN:AAA52370.1; PID:g553272
 R:Halley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.
 Oncogene Res. 1, 375-396, 1981
 A:Title: The human EGF receptor gene: structure of the 110 kb locus and identificat
 A:Reference number: S30024; MUID:88217333; PMID:3329716
 A:Accession: S30024
 A:Molecule type: DNA
 A:Residues: 1-29 <HA2>
 A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
 R:Halley, J.D.; Waterfield, M.D.
 J. Biol. Chem. 266, 1746-1753, 1991
 A:Title: Contributory effects of de Novo transcription and premature transcript tex
 A:Reference number: A38672; MUID:91107677; PMID:1988448
 A:Accession: A38672
 A:Molecule type: DNA
 A:Residues: 1-29 <HNL>
 A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271
 A:Experimental source: carcinoma cell line A431-7
 R:Xu, Y.; Ishii, S.; Clark, A.J.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.
 Nature 309, 806-810, 1984
 A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of
 A:Reference number: A00642; MUID:84245835; PMID:6330563
 A:Accession: A00642
 A:Molecule type: mRNA
 A:Residues: 150-187, 'KSVIGAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 30
 '798-799, 'TP', 802-811, 'R', 813-942 <XUV>
 A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF
 R:Lin, C.R.; Chen, W.S.; Krutger, W.; Stolarsky, L.S.; Weber, M.; Evans, R.M.; Ver
 Science 224, 843-848, 1984
 A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica
 A:Reference number: A43615; MUID:84196372; PMID:6326261
 A:Accession: A43615
 A:Molecule type: mRNA
 A:Residues: 713-964 <LIN>
 A:Experimental source: epidermoid carcinoma cell line A431
 R:Stamen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984
 A:Reference number: A23062; MUID:85046483; PMID:6093780
 A:Accession: A23062
 A:Molecule type: mRNA
 A:Residues: 1028-1210 <SIM>
 R:Weber, M.; Gill, G.N.; Speiser, J.
 Science 224, 294-297, 1984
 A:Reference number: A05281; MUID:84172183; PMID:6324343
 A:Accession: A05281
 A:Molecule type: protein
 A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
 R:Russo, M.M.; Lukac, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A:Title: Identification of residues in the nucleotide binding site of the epidermal
 A:Reference number: A60143; MUID:85182650; PMID:2985980
 A:Accession: A60143
 A:Molecule type: protein
 A:Residues: 740-744, 'X', 746-747 <RUS>
 R:Moczekowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and su
 A:Reference number: A38023; MUID:84191554; PMID:6325948
 A:Contents: annotation; receptor activity

A>Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsch, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A>Title: Functional independence of the epidermal growth factor receptor from a domain
 A:Reference number: A33183; MUID:900032233; PMID:2790960
 A:Content: annotation; internalization signal
 C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
 C:Genetics:
 A:Gene: GDB:EGFR
 A:Cross-references: GDB:120610; OMIM:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:124/DNA: signal sequence #status predicted <SIG>
 F:124/DNA: Product: EGF receptor #status predicted <MAP>
 F:25-1210/Product: EGF receptor #status predicted <EXT>
 F:25-645/DNA: extracellular #status predicted <EXT>
 F:75-300/DNA: EGF receptor extracellular domain repeat <E21>
 F:390-600/DNA: EGF receptor extracellular domain repeat <E22>
 F:669-1210/DNA: transmembrane #status predicted <TM>
 F:669-1210/DNA: intracellular #status predicted <INT>
 F:710-975/DNA: protein kinase homology <KIN>
 F:718-926/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128-175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
 F:745/Active site: Lys #status experimental

Query Match 45.6%; Score 3113; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 6,3e-121;
 Matches 625; Conservative 173; Mismatches 363; Indels 102; Gaps 21;

```

11 LLLALLPFGAA--STQVCTGDMKLLPASPETHLDMKRLHYGCGVQGNLETLPLPN 68
14 LLALPASPALFEKKVCCQSTNKLTLQGFEDHFLSLQFNNCCVVGNEITVQEN 73
69 ASLSFLQDIOEVGVYLAHQVRYQLRLIRIVRGTLFEDVYALAVLDNGDPLNTP 128
74 YDSFLKTIQEVAGVYLAHQVRYQLRLIRIVRGTLFEDVYALAVLDNGDPLNTP 126
129 VTASAPGGLRELOLSLTLLKGVLIQRNPOLCYDITLMDIPIKNNQALITLIDNR 188
127 ---ANKTGLKELMRNLOELTHGAVFNNPFCANVESIQMDIVSDFLSNMSMDFOHN 183
189 SRACHPCSPWCKSGRCSCGSESSDQSLTRIVCGAGA-RCKGFLPTDCCHECAACTCP 247
184 LGSCKQKDBSCPCSGSCWAGENCQKTLKICQCSGRCRGSBPCCHQCAAGCTGP 243
248 KHSDDLACLFHNSGICELHCPALVTYNTDTPESMDNPGRYTFGASCVTACPYNYLSTD 307
244 RESDCLVCRKFRDEATCKDTCPLMCTYNTQMDVNPBGKYSFGATCVKCKPRNVVTD 303
308 VGSCTLVCPPLHNGEVTAEDGTORCEKSCPCARVCYGLQMOYIKANSKFIQTELE-FAG 366
304 HSGCVACAGADSTEM-EDGCVKCKCKGECRCRVCGIGGEFR-DLSINATNTKHFKN 361
367 CKKIFGSLFLPESPDGDPASNTAPLQPEOLQVFEITLLEITLGYLISAMPDLSPLSVQ 426
362 CTISGDLHLPLVAFRGDSFTTHPPDQGLDILKTVKEITGLLQAMPENRTDHAHE 421
427 NIQVITGRILLHNAVYLLTQGLISWLGRLSLRELSSGLALHNTLHLCFVHTVPDOLF 486
422 NNEITGRITKQHQFLAVASLNTSLGRSLKEISDGVIIISGNKLCYANTIMWKLF 481
487 RNHQALHTANREDECVGEGLACQLCARGCHWCPPTQCNQCSQFLRGQECVEGV 546
482 GNSGCKTKLISNNGENSKATQGVCAALSPEGCWCPRPDVCSCNVSRGSECDKCKL 541
547 LQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAPCSGYKP 606
542 LEEGPREFVENSECICHCECLFOANNITCTGKGPNCITQCAHYIDGPHCVCTCAGVWG 601
607 DLSYPMKFPDEGACQCPICNCTHSQVLDLQKGPAPGRASPLTISNTNTVFWLRLV 666

```

Db 602 ENNTL-VKTYADAGVCHLCHPNCYGGCTGPGLESCPTNGKIP--SIATGVNALLLL 658
 QY 667 PVASASHLEKROQKIRKYTWRLLOEELVEPLTPSGAMPYQAOMRLIKENELAKVYL 726
 Db 659 VVALGILPMRRRIYVKRKLRLRLQELVELPLPSGAPALRLILKTEFEFKIKYL 718
 QY 727 GSGAGTYVKGIMPIDGENVKIPVAKYLRNTPSKANKELIDEAYVAVGSPVYSRL 786
 Db 719 GSGAGTYVKGIMPIDGENVKIPVAKYLRNTPSKANKELIDEAYVAVGSPVYSRL 778
 QY 787 GICLSTVQLTQMLPFGCLLDHYENNGRLGSDLLWCMQIANGSYLDEVLYRDL 846
 Db 779 GICLSTVQLTQMLPFGCLLDHYENNGRLGSDLLWCMQIANGSYLDEVLYRDL 838
 QY 847 AARNVLYSPNHKKTIDFGARLDIDREYHADGKVPKMMALLESILRRFTQSDVW 906
 Db 839 AARNVLYTPQHVKTIDFGARLDIDREYHADGKVPKMMALLESILRRFTQSDVW 898
 QY 907 SYGVTVWELMTFGAKPYDGIAPARLIDLEKGERLPQPPCTIDVYIMVCMIDSECR 966
 Db 899 SYGVTVWELMTFGAKPYDGIAPARLIDLEKGERLPQPPCTIDVYIMVCMIDSECR 958
 QY 967 PRFRLEVEESFMAADPQRFVIO-NEDLGSASPDSFFYSLSLEDDMDGLVDAEYLV 1025
 Db 959 PRFRLEVEESFMAADPQRFVIO-NEDLGSASPDSFFYSLSLEDDMDGLVDAEYLV 1018
 QY 1026 PQGFCFCDPAPGAGMWHHRSSSTRSGGADLLTGLPESEEARSPAPSEAGSDV 1085
 Db 1019 PQGFCF-----SSPSTRPFLSLSATSN- 1043
 QY 1086 PDGDLGMAAKGLQSLPETHDSPLQRYSEDTPVLPST--DGVAFLTCSPQPYVNP 1143
 Db 1044 -NSTVACIDRNGLOSPCKEDSFLQRYSSDTGALTEDSIDTDFL-----PVSEYINQ- 1095
 QY 1144 DVREPQPPREGPPLAARPACTLEAKTSLSPKGVYKDVFAFGAVENPEV-TPQG 1202
 Db 1096 SVKPRPASYQNPVTHNPVLP-----APSRDPIQD--PHSTAVANGELYNTVO-- 1143
 QY 1203 AAPQPPAPSPAFDNLVYDQ-----DP-----PERGAPSTFKGPTAENP 1246
 Db 1144 -----PTCVNSTPDSFAHWAQKSHQISLDNPDYQDFFPKKAKPRGIFKGS-TAENA 1195
 QY 1247 EYL 1249
 Db 1196 EYL 1198

RESULT 5
 A33183
 A:Title: epidermal growth factor receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C:Accession: A33183; A33181; S24942; A28941; S45325; 19643
 R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N
 Genes Dev. 9, 399-413, 1994
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF recep
 A:Reference number: A33183; MUID:94170986; PMID:8125255
 A:Accession: A33183
 A:Molecule type: mRNA
 A:Residues: 1-1210 <LUE>
 A:Cross-references: GB:U03425
 R:Aviyl, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
 Oncogene 6, 673-676, 1991
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding
 A:Reference number: A3318; MUID:91232866; PMID:2030916
 A:Accession: A3318
 A:Molecule type: mRNA
 A:Residues: 1-714 <AVI>
 A:Cross-references: GB:X59698
 R:Elstinger, D.P.; Serrero, G.
 Submitted to the EMBL Data Library, June 1992
 A:Reference number: S24942
 A:Accession: S24942

A:Molecule type: mRNA
 A:Residues: 969-971,'K',973-1115,'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heistermann, G.U.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13159, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:8830814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: Protein
 A:Residues: 689-694,'X',696-704,'L',706-707,989-992,'XX',995-996,'X',998-1000;1002-1009,
 R:Hisbbs, M.L.; Dunn, A.R.; Alexander, W.S.
 Submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: 945325
 A:Accession: 945325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971,'K',973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CAA55587.1; PID:9488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
 A:Reference number: 149643; MUID:93126380; PMID:7678348
 A:Accession: 149643
 A:Status: translated from GB/EMBL/DD81
 A:Molecule type: mRNA
 A:Residues: 12-20,22-132 <RES>
 A:Cross-references: GB:U06864; NID:9193001; PIDN:AAA53029.1; PID:9567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680-695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:687,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.1%; Score 3080; DB 2; Length 1210;
 Best Local Similarity 49.3%; Pred. No. 1.4e-119;
 Matches 627; Conservative 164; Mismatches 371; Indels 110; Gaps 23;

11 LLLALLPPGAA--STGYCTGTDMKRLPASPEHLDMLRLAYOGCQVQVQCNLELYLPTN 68
 14 LTRALCAAGALEKKKVCQGTNRRLTQGTFFEDHFLSLQRMYNNEVGLNLEITYVQNR 73
 69 ASISFLDIOEVQGVYLAHNOVRQVPLQRLRYRQQLFEDNYALAVDNGPLNNTTP 128
 74 YDSFLKTIQEVAVYLAHNTVERLPLENLQIIRGALYENTYALALISN----- 124
 129 VTGASPGGLRELQRLSTLEILKGVLLIQRPOLCYQDTILMKDI---FHKNNQLALTLI 184
 125 -YGNTRGLRELPRNLOETILIGAVRSPNNPILCNMDTIQMRDIYQNVFNSNMDL---- 180
 185 DTRRSRACHPCSPKCKSRGCGESSSDCQLTATVACGGA-RCKGPLPTDCCHEQCAAG 243
 181 -QSHPSCCPKCDPCSPGSCGSGGEEQCRLTKITCAQCSHRCRGSPSDCCNOCAG 239
 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRYTFGASCTACPVNY 303
 240 CTGPRESDCLVCKQFOEATCKDTCEPLMLYNTTYQMDVNPFGKRSFGATCYCKKPRNY 299
 304 LSTDVSGCLVCPHNOVTABDGTQRCCKSPCARCYGLQMYQYIKANSKFIITLE 363
 300 VVVDHSGCVACGDDYEV-EEGQIRKCKKCDPCPKVCGIGICGFEK-DTISINATNIK 357
 364 -FAGCKKIFGSLAFPLPSFGDDPASNTAPLOPEOLQVETLEETITVLYLSAMPDLPDL 422
 358 HFYVCTAISGDHLHLPFAFKGDSFTRPPLDFRELEILTKVKEITGFLLIQAMPDWTDL 417

423 SVFQNLQVIRGRILHNGAVSLTLOGISWLGSRSLGSCALIHNTLCFVHTVPM 482
 418 HAFENLEIRGRTHQGFSLAVGLNTSLGRLKEISQDDVITSGNRNLCYANTINM 477
 483 DOLFNNPQALHTANREDECEVGEGLACHOLCARGHMGPGPTQVACNSOFLRQCEVE 542
 478 KULFGTPQKTKIMNNRABKCKKANVHCNPLCSSECGMGEPKRCVACNSRRECEVE 537
 543 ECRVTLQGLPREYVNAHCLPCHPECPQNGSVTCGEPADQVACAHKDPFCVACRPS 602
 538 KNILGEGEREFVENSECICQHPCECLPQAMNITCTGRSPDMCICQAHYIDGHCYKTPA 597
 603 GYKPLSYVPIKPFDEBACOPCPINCTHSCVDLDDGCPAECQASPLSTIFNFTVSF 662
 598 GIMGENNLT-VKXVDANNVCHLCHANCTYGCAGGLOGCEVWSPGKPIPSIATQVGL 656
 663 WLRVPSASHLEKRRQKIRKYTRMRLLQETELVEPLTPSGAMNQAMRLKETELRK 722
 657 LFIIVVAALGIGLFMRRIHVRKTLRLLQERELVEPLTPSGEAPNQALRLKETEFKK 716
 723 VLVLSGARGTYKGIWIPDGENYKIPAIKVLRENTSPKANKEILDEAVYVAGVSPYV 782
 717 IKVLDSGARGTYKGIWIPDGENYKIPAIKVLRENTSPKANKEILDEAVYVAGVSPYV 776
 783 SRLIGCLSTVQVLTQMLPYGCLLDHYENRGRGLSODLLNMCQIAKMSYLEVRLV 842
 777 CRLIGCLSTVQVLTQMLPYGCLLDHYENRGRGLSODLLNMCQIAKMSYLEVRLV 836
 843 HEDLAARNVLYKSPHNVKTTDFGLARLLDIDETEHADGKVPYIMWALESTLRFRFHQ 902
 837 HEDLAARNVLYKSPHNVKTTDFGLARLLDIDETEHADGKVPYIMWALESTLRFRFHQ 896
 903 SPVMSYGVWELMTFGAKPYDGIPIAREIPDLKEGELPOPICTIDVYIMVYCMWID 962
 897 SPVMSYGVWELMTFGAKPYDGIPIAREIPDLKEGELPOPICTIDVYIMVYCMWID 956
 963 SECRPRPELVSEFSRMARDPQRFVIO-NEDLGASPLDSTFYKSLLEDMDGLVDAE 1021
 957 ADSRKPEFLLEIFGQMRDQRYLVICGSRMHLPSPDSENFYALDEEDMEDVAD 1016
 1022 EYLVPQGFPCDPAKAGKGMVHHRSSFRSGGDLTLGLEBEEAPSPSLAPSSGA 1081
 1017 EYLVPQGFPCDPAKAGKGMVHHRSSFRSGGDLTLGLEBEEAPSPSLAPSSGA 1042
 1082 GSDVFDGDLGMAKGLSLPTHPSPQCRSEDEPTVLPSET--DGVAALTCSPQPEY 1139
 1043 TSN-----NSTVACINRNSCRVKEDAFIQRYSSDPTGAVTEDNIDDAFL-----PVPEY 1092
 1140 VNQDPVYRQPPSPREGPLPAARPAATLEAKTISPGKNGYKDYVAFAGAVENPEYL-T 1198
 1093 VNQ-SVPRPAGSVNPNYHNPPLH-----APGDLHYON--PHSNVAGVEYNT 1141
 1159 PQGAPAPRHPPPAPFNDLXYWDQ-----DP-----PERGAPSPSPKPTPT 1242
 1142 AQ-----PTCASSGFNSPALMIGQSHQMSLDNDYQDDFFPKTKTNGIFXG-PT 1191
 1243 AENPEYGLDVP 1254
 1192 AENPEYGLDVP 1203

RESULT 6
 TCVHLY
 epidermal growth factor receptor precursor - chicken
 N:contains: protein-tyrosine kinase (EC 2.7.1.112) erdb
 C:species: Gallus gallus (chicken)
 C:date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C:Accession: A27720; A00643
 Relax: I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Ver
 Mol. Cell. Biol. 8, 1970-1978, 1988
 A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in
 A:Reference number: A27720; MUID:88261272; PMID:3260329

A/Accession: A27720
 A/Molecule type: mRNA
 A/Residues: 1-1223 <LAX>
 A/Cross-references: GB:M20386
 R/Nilsen, T.W.; Maione, P.A.; Goodwin, R.G.; Rotman, F.M.; Ciltenden, L.B.; Raines, M.
 Cell 41, 719-726, 1995
 A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A/Reference number: A00643; MUID:85228222; PMID:2988784
 A/Accession: A00643
 A/Molecule type: mRNA
 A/Residues: 585-1223 <NIL>
 A/Cross-references: GB:M10066
 C/Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
 C:KeyWords: epidermal growth factor receptor #status predicted <SIG>
 F:1-30/Domain: signal sequence #status predicted <SIG>
 F:1-1223/Product: epidermal growth factor receptor #status predicted <MAT>
 F:1-654/Domain: extracellular #status predicted <EXT>
 F:61-307/Domain: EGF receptor extracellular domain repeat <E1>
 F:397-610/Domain: EGF receptor extracellular domain repeat <E2>
 F:655-677/Domain: transmembrane #status predicted <TM>
 F:678-1223/Domain: intracellular #status predicted <INT>
 F:719-984/Domain: protein kinase homology <KIN>
 F:727-735/Region: protein kinase ATP-binding motif
 F:136-202,280,361,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #
 F:192,650/Binding site: carboxylate (Ser) (covalent) #status predicted
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:754/Active site: Lys #status predicted
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.6%; Score 3049.5; DB 1; Length 1223;
 Best Local Similarity 48.3%; Pred. No. 2,66-118;
 Matches 626; Conservative 166; Mismatches 360; Indels 145; Gaps 25;

8 RMGLLALLPAGA-----STQVCTGTMTKLRLPASPETHLDMRLHYCGQVQGNLE 61
 13 RGAVALVLLGLGALCSAVEBEKKVCGGTNNKLTQGHVDHFTSLQRMNNECVLSNLE 72
 62 LTYLPNALSFLQDQVQGYVLTANQVROVPLRLIIVGTQIFEDNALAVLDNGD 121
 73 IYVEHNRDLTLKTOEAVGYVLTANNDVPIPLENIOIIRGNVLYDNSEFALAVSNH 132
 122 PNNTPPTVAGSPGRLREQLRSLEILKGVLIQNPQLQVODIIMKIDHKNQAL 181
 133 -NKKQ-----GLRELPMKRLISILNGSVKISNNPKLCNMDVLMNDIIDSRL-PL 182
 182 TLID-TNRSRACHPSGKSGSRCWSESEDCQSLTRVCAGGA-RCCKPLPTDCHEQ 239
 183 TVLDFASNLSCPKCHPCTEDHGWAGEQNCQTLTKVCAQCCSGRCGRKVPSCCHNQ 242
 240 CAAGCTGPKGSDCLCLHNHSGICELHCPALVTYNTDFESMPREGRYTFGASCYAC 299
 243 CAAGCTGPKGSDCLCLCRKRRDATCTCCTPLVLTNPTTYQNDVPEGRYSFGATCYEC 302
 300 PNYLSTDVSGCTLVCPFLNDEVTADGTQCEKSKPCARCYCLGMQYIKANSKFGIGI 359
 303 PHNYVYTDHSCVRSQNTDTEYV-BENGVKCKCKDGCLCKKCNIGIGELKGLIS-INA 360
 360 TLEL-FAGCKKIFGSLAPLPESFPDDPASNTAPLPDQIQVETLEBITTYIYISAMPDS 418
 361 TUIDSPKCTKNGVSLIPVAFIDAFKTLPLDPKLDVFRYKEISGFLIQAMPDN 420
 419 LPDLVFOQLQYIRIRLHNGAVSLTLOGLISMTGLSLRELSGSLLIHNNTHLCVH 478
 421 AFDVAFENLEIRRTKHQGOYSLAVNLKQISGLSKLEISGDIAMKNKMLCTAD 480
 479 TVPMDLPNPHQALLHNPREDCEVGBGLACHQLCARGHGWPFGPTQVCNCSQGLRGQ 538
 481 TNNRSLFATOSQKTKIIONKNDCTADRHVCDFLCSDVGCMGPGFPHCSCPFFSRQK 540
 539 ECVEECRVLQGLPFEYVNAHCLPCHPEQONG---SVTGGPEADCCVACAHYKDPFF 595

541 ECVKQCNILQGEPRFERDSKCLPCHSCLVONSTAYNTTSCSPGPDHCKMCAHFDGPH 600
 596 CVARPSGVKPDLSMTWPKFPDEGACQPPCPINTHSCVDLDKGGCAEGRASLTSTF 655
 601 CVKAPAGVLENDTL-VKTDANAVQLCHPNTCTCKCGGLECCP---NGSTPSTIA 656
 656 NNFVTSFWLRVPKVSASHLEKROQKIRKRYTMRRLTCELEVEPTPGAMPNQOMRIL 715
 657 AGVGGGLCLVVGIGILYLRRIYAKRLTLRLQRELVEPTPGAMPNQOMRIL 716
 716 KETELRKXVVGSGAFGVYKGMIPDGENYKIPVATVLRBNTSPKANKELDEAYMA 775
 717 KETEFKXVVGSGAFGVYKGMIPDEGKXIPVATVLRBNTSPKANKELDEAYMA 776
 776 GVSPPVYRLIGICTSTVQVLTQMPYGLLDHREBRGLSDGLNMQOAKGMSY 835
 777 SVDNHRVRLIGICTSTVQVLTQMPYGLLDHREBRGLSDGLNMQOAKGMSY 836
 836 LEDVLRDLAARNVLYKSPBNHYKIDFGRLALDIDETEVHADGCKVPIKMALESIL 895
 837 LEERLVARDLAARNVLYKTPQHVKITDFGLAKLLGADKEKYHAEGKVPIKMALESIL 896
 896 RRRFTQSDVMSYGVTWELMTFGAKRPDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955
 897 HRIYHSDVMSYGVTWELMTFGSKPDGIPASEISVLEKGERLPQPICTIDVYIM 956
 956 VKCMWIDSECPREFELVSEFSRMAPQRFVVIQ-NEDLPASPLDSTFRSLDEDDM 1014
 957 VKCMWIDSDSPKRELIAESVKAARDPRRLVIOGBERHMLPPTSKFRITLMEEDM 1016
 1015 GDLYDAEYLYPQCGFPCPDPAAGAKVHHRRSSSTRSGGDLTGLEPSEEARSP 1074
 1017 EDYDAEYLYPQCGF-----NSPST-----SRTP 1042
 1075 L-----APSEGAGDVDFGDLGMAAGLQSLPHDPSPIQRYSEDETPVLPST--DGY 1127
 1043 LLSLSLSTSNNSATNCID-----RNGCHPREDSEFVQRYSDPFGNLESDIGF 1094
 1128 VAPLTGSPQEPYVNPDPVPPSPRPGPLPAPRAGATLERAKTLSPGKXGVKDYVF-- 1185
 1095 L-----PAREYVQ--LMPKPS-----TAMVQNOIYNNISLT 1125
 1186 -----AFGAVENPEYLTQGGAAAPPHPPAPSPAFDNLVYWDQ----- 1225
 1126 AISKLPMSRYQNSHTAVNPEYL-----NTNOSPAAKTVESSPYWIOGNHQN 1177
 1226 -DPE-----RGAPSTFGKPTAENPEYLGIDVP 1254
 1178 LDNPDYQDDPLNPETKENGLLKVAENPEYLRVAP 1214

RESULT 7
 A47253
 epidermal growth factor receptor, HER4 - human
 C/Species: Homo sapiens (man)
 C/Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
 C/Accession: A47253
 R/Floman, G.D.; Culicou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
 A/Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epider
 A/Reference number: A47253; MUID:93189574; PMID:833326
 A/Accession: A47253
 A/Status: preliminary; not compared with conceptual translation
 A/Molecule type: nucleic acid
 A/Residues: 1-1308 <ELO>
 A/Cross-references: GB:L07868; NID:9337359; PIDN:AAB59446.1; PID:9337360
 A/Note: Sequence extracted from NCBI backbone (NCBI:P126842)
 C/Suprafamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; growth factor receptor
 F:716-981/Domain: protein kinase ATP-binding motif
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 42.7%; Score 2920.5; DB 2; Length 1308;
 Best Local Similarity 44.6%; Pred. No. 5.5e-113;
 Matches 604; Conservative 178; Mismatches 384; Indels 187; Gaps 26;

9 MGILLALLPPGAA---STQVCTGTDMKRLPASPETHLDMRLHLVQGCQVQGNILETV 64
 8 MWWWLLVAGTVQPSDQSVCACTENKLSLSLEQYRALRKYENEVEVMGNLEITS 67
 65 LPTNASLSPDIOIEVQGYVLIANQVROPLORLIVRGTOLEFEDNYALAVDNGDPLN 124
 68 IEHNDLSFLRSYREYVGYVALNQFRYLPENELIRGTXYERVALAFYLRKNDG 127
 125 NTPPTVAGSPGRLREIQLRSLEILKGVLIQNPOLCYODTILMDIETHKNNQALTLI 184
 128 NF-----GLQELKMLTEILNGVYVNDKRLCYADTIHMDIYRANRPSLTIV 178
 185 DTRSPACHPSCPMKSGRCWGSSEDCSLTETVCAAGC-ARCKPRLPTDCHQEOCAAG 243
 179 STNGSGGGRCHKSCG-RCWGPTEHNCQTLRTVCAEQDCRCYPRVSDCHRECAAG 237
 244 CTGPKSDCLACLFHNSGICELHCPALVYNTDTESMNPREGRTFGASCYTACPVNY 303
 238 CSGKPDVCFACMNFDSGACVTCQCPTEFVYNTTQLHEHNAKTYGAFVCKCPHNF 297
 304 LSTDVGSCTLVCEPLHNOEYTABDQRCCKSKPCARVCYGLMOKYIKANSKFTIGTELE 363
 298 V-VDSSSCVACPSKMEV-BENGIKCKRPTDIOCPACGIGTGSLSMAQTVDSNIDX 355
 364 PACCKTFGLATLPESFDDPASNTPLOPQOLVETLEITGYLIYSAMPDLSPLDS 423
 356 FINCTKINGNLIPLVTHGDVNAIEADPEKLNFRVREITGFLNTQSPMPNTDS 415
 424 VEQNLQVIRRIILHNGAYSLTLOGLISMLGSLRLPSLGLALIHNTLCEVAFVPM 483
 416 VFSNLVITGRVLYSGSLILKQGLTSLQFOSLKSIGNIYITDNSLVCYHINMT 475
 484 QLEFNPHQALLHTANPEDECVGEGLCACHQLCARGHCWGPFPQVCNCSOLFQCEVEE 543
 476 TLFSTINQIVIRDNRAKENTAGMVCNHLCSDDGCMGPGPOCCSCRFSGRITCIES 535
 544 CRVYQGLPREYVNAHRLPCHPEGOP-QNGSVTCFGEADQCAVACHVYKPPRCVACRS 602
 536 CNLDGFFRENGSICVECDPOCEKMEGDLTCHGPGPNCTKCHFKGRCVCEKCPD 555
 603 GVKEDLSYMPWKEPDEGACQPCFINCTHSCVDLDCGPABORASPLTSLF---NNF 658
 596 GLQGANSP--IFKYADPDRECHPCHPCTOCNGPTSHDC-----IYPTWGHG 642
 659 TVSFMLRVPKVSAS-----HLEKRRQCKIKKTYMRRLLOTELVETPLPS 703
 643 TLPGHARTPLIAAGVIGGLFLVIVGLTFPAVYRRKSIKKRALRRL-ETELVEPLTPS 701
 704 GAMPNQAOMRLKETELRYKVLGSGAFIVYKGIWIPDENVKIPIAKIVLRBENTSFK 763
 702 GTANQAOLRLKTELKRYKVLGSGAFIVYKGIWIPDEBYTKIPIAKIILMETTPKA 761
 764 NKEILDAVYVAGSGPYVSRFLGICITSTVQVLTQMPYGLCLDHYREKRGKSGODLL 823
 762 NVEFMDEALIMASMDHHLVRLGVCCLSPITQVLTQIMPGCLLEYVHEKMDINGSL 821
 824 NMCOIAKMSYLEDLVLRDLAARNVLVKS PMVHKITFGFLARLDLIDETEYHADGK 883
 822 NMCOIAKMSYLEDLVLRDLAARNVLVKS PMVHKITFGFLARLDLIDETEYHADGK 881
 884 VPIKMALESTILRRFTHQSDVSYGVYVWELMTFGAKPYDGIPTREIPLLEKGERLPQ 943
 882 MPIKMALESTILRRFTHQSDVSYGVYVWELMTFGAKPYDGIPTREIPLLEKGERLPQ 941
 944 PICTIDVYIMVYCMWIDSECRPRFELVSEFSRMARDPORFVIONED-LGPASPLDS 1002
 942 PICTIDVYIMVYCMWIDSECRPRFELVSEFSRMARDPORFVIONED-LGPASPLDS 1001
 1003 TFRSLLEDDMDGLVDAEYLVQGFPCDPAPAGAGVHHRHRSSTRSGGDLTLG 1062

1002 KFPQNLDEEDLEMDMAEYLV-QAENIPEP-----IYTSARIDSNS-----EIG 1049
 1063 LEPEEEAPRS-----PLAP-SEGAGSVFPGDLCMG 1093
 1050 HSPPAVYPMGSGNFVYRDGGAFAEGSVYFAPRTSTIPAPVQAQATABI FDDSCNG 1109
 1094 AAKGQSLPTHDSPRLRYSEDPVPLPS-----ENDGVAPLTCSPQEEYVNOPDV 1146
 1110 TLKRFVAHVVEDSTQRYADTVFAPERSPRELDEEGYMTMRDKPKOBYLNPVE-- 1167
 1147 PQPSFREGPLPAARPAQATLERAKTISPGKNGVYKQVFAFGAVENPEYLTPOGGAAPQ 1206
 1168 -----ENPVSRR-----KNGDLQ-----ALDNEFYHNASNG----- 1194
 1207 PHPPA-----FSPAFDNLVYWDQPPPERGA 1232
 1195 --PPRAEDVYNEPLVNTFANTLGAKEYLNNILSMPEKAKKAFDPDYNHSLPFRST 1252
 1233 --PSTFKGTPT-----AENPEYL 1249
 1253 LQHPDYLOEYSTKYFYKNGRIRPIVVENPEYL 1295

RESULT 8
 506142
 protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
 N.Alternate names: epidermal growth factor receptor homolog; kinase-related transfo
 C.Species: Xiphophorus maculatus (southern platyfish)
 C.Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
 C.Accession: S06142; S13809
 C.Mitochondr: J.; Adam, D.; Maltischek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Ro
 Nature 341, 415-421, 1989
 A.Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
 A.Reference number: S06142; MUID:90015140; PMID:2791166
 A.Accession: S06142
 A.Molecule type: DNA
 A.Residues: 1-1166 <WTT>
 A.Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
 R.Adam, D.; Maeueller, W.; Scharf, M.
 C.Jogene 6, 73-80, 1991
 A.Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
 A.Reference number: S13807; MUID:91125582; PMID:1846957
 A.Accession: S13809
 A.Status: preliminary; translation not shown
 A.Molecule type: DNA
 A.Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
 A.Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
 C.Genetics:
 A.Gene: mrk
 A.Map position: Y
 A.Introns: 872/3; 988/1; 947/1; 979/3; 1025/3; 1056/1
 C.Superfamily: epidermal growth factor receptor; protein kinase homology
 C.Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
 F.1-25/Domain: signal sequence #status predicted <SIG>
 F.26-1166/Product: kinase-related transforming protein (Tn) #status predicted <MAT>
 F.707-972/Domain: protein kinase homology <KIN>
 F.715-723/Region: protein kinase ATP-binding motif

Query Match 38.4%; Score 2627; DB 1; Length 1166;
 Best Local Similarity 44.7%; Pred. No. 5.7e-101;
 Matches 568; Conservative 165; Mismatches 393; Indels 144; Gaps 31;

4 AALCRGILLALLPPGAASST---QVCTGTDMKRLPASPETHLDMRLHLVQGCQVQGN 59
 8 AALIQ--LTLVLSISRCSTDPDRKVCQGSTNQMT---LDNHYLRMKMVSQGVAVLEN 62
 60 LEITYLPTNASLSPDIOIEVQGYVLIANQVROPLORLIVRGTOLEFEDNYALAVDN 119
 63 LEITYLPTNASLSPDIOIEVQGYVLIANQVROPLORLIVRGTOLEFEDNYALAVDN 122
 120 GDLNNTPTVAGSPGRLREIQLRSLEILKGVLIQNPOLCYODTILMDIETHKNNQAL 179

Db 123 YOK-NPSSP--DVTQVGLKQJOLSLNLEITLSSGKVSNNPLLCVETINMMDIVDKTSNP 179
 QY 180 ALTLIDITNRSPACHPCSPCKGSRGSSSEDCSLRTVAGGC-ARCKGPLPTDCHE 238
 Db 180 TNNILPHAFERQCCQCHGCVNGSCWAPGHCQFTLLCAECNCRCKRPKIDCNE 239
 QY 229 QCAAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTESNPNRPGRTTPASCVTA 298
 Db 240 HCAAGCTGPRATDCLACRDFNDGCKDTCPPKIDYIVSHQVVDNPNKITYFPAACVKE 299
 QY 299 CPYNYLSTDVSGSCTVCPHNOEYTAEDGTORCEKSPCARVCGQMGQYIKNSKFIG 358
 Db 300 CPNSVYVTE-GACVRSACAGLBYD-ENGKSKCPKCDQVCKVCDGIGISGL-SNTIAVN 356
 QY 359 ITEL-EPAGCKKIFGSLAF.PESPDGDPASNTAPLOPQLOVFETLEITGYLYISAMPD 417
 Db 357 STNIRSFENCKTINDIILNHNSEGDHPYKIGTMDPEHNLNLTVEITGYLYIWMWPE 416
 QY 418 SLPLSVFQNTQVIRGRILHNGAYS-LTLQGLG:SWGLRSLRSLGSLALIHNTLCLF 476
 Db 417 NMTLSVQNLLEIRGRITTSRGSFVAVQVRHQLQWGLRSLKESAGNVILKNTLQLRY 476
 QY 477 VHTVWQDLFRNPQALHTANRDEDECVEGLACHQ:CARHGMGPGPTQCVNCSQFLR 536
 Db 477 ANTIMRRLFRSEDOSEIYDART-----ENQICNNECSEDCGM-PGPTWCVSCLAHYDR 528
 QY 537 GQECVEGRVLQGLPREYVNAHCLPCHPEQOPONGSVTCGPRADQVACAYKDPFPC 596
 Db 529 GGRCVASCNLLQGEFREAOVDGRQVQOECLVQDTSITCGPBPANCSKSAHFDPOC 588
 QY 597 VAPCPGVKPDLSWPIKFPDEGACQPCINCHSCVDDDDKGPAAE-QRASPLTSI 654
 Db 589 IPRCHGILGSDTL-IMKADKMGQCPCHQNTQCGSGGLSGCAGDIYSHSLAVGL 647
 QY 655 FNNFTVSFWLRVPKVSASHLEKRPQOKIKRYTMRLIQETELVEPLTPSGAMPQAMRI 714
 Db 648 VSGLLITVALLIV--LRRRRRIK-RKRTIRCLQEKELVPLTPSGQAPQALRI 703
 QY 715 LKETELRVKVLGSGAGTGYKGIWIPDEGVNKPVAKYLRKENTSKAKELIDEAYVM 774
 Db 704 LKETEFKDRVLGSGAGTGYKGLMNDGENTIRPVAKYLRKENTSKAKVOEVIDEAYVM 763
 QY 775 AGVGSFYVSLGLCLSTVQVLTQLMPYGLDHNENRGLSQDLNWMQIAKMS 834
 Db 764 ASVDHPHYCRLLGICLSAVQVLTQLMPYGLDLYVAHQERICQGLNWMQVQIAKMN 823
 QY 835 YLEBVLVHRDLAARNLVKSPNPKITDTEGLALDIDETIYHAGGKVPKIMMALESI 894
 Db 824 YLEBRHLVHRDLAARNLVKSPNPKITDTEGLALDIDETIYHAGGKVPKIMMALESI 883
 QY 895 LRREFTHQSDVWSYGVTVWELMTGKAPYDGIIPARETIPDLKXGERLPPOPICITIDYMI 954
 Db 884 LQWYTHQSDVWSYGVTVWELMTGSKPYDGIIPAKELASVLENGERLPPOPICITIEYMI 943
 QY 955 MYKCMWIDSECRPFRELVSEFSRMAADPQRFVVIQNEIDLGAPSLDSTFYRSLLEDDDM 1014
 Db 944 ILKCMWIDPSRPFRELVSEFSRMAADPQRFVVIQNEIDLGAPSLDSTFYRSLLEDDDM 999
 QY 1015 GDLYDAEBEYLVPQGGFCPPRPARAGGMVHHRHSSSTRSGGGLITGLERSEEARSP 1074
 Db 1000 -DAVDADEYLLPYKRI-----NRQGS-----EP 1021
 QY 1075 LAPSEAGSDVFDGLMGAKGLQSLPTHDPSPLQYSSEDPV-PLPSETDYVAFLTC 1133
 Db 1022 CIPPTGH-----PVRENSITLANSIDPQNALKLODGH----- 1055
 QY 1134 SPQREYVNOFVRFQF-----PSPRE-----GPLP-AARPGATILERAKTSLPGKNGV 1180
 Db 1056 ----EYVNOFSESTSSRLSDIYNNENYEDLIDGMQPVSLSSQEAETNSRREYVNTQNSL 1111
 QY 1181 VKDVFAGGAVENNEYETTPQGAAPRPHPPAPSPADNLYWDDPPEPGAGAPSPFFKGT 1240
 Db 1112 ----PLVSSGMDPBDY---QAG-----YQAA-----LPQGLALTGGMFL 1146

QY 1241 PTANPEYLG 1250
 Db 1147 PAENLEYLG 1156
 RESULT 9
 A36223
 kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
 C:Accession: A36223; 159164
 R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
 Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
 A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epide:
 A:Reference number: A36223; PMID:90083234; PMID:2687875
 A:Accession: A36223
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1342 <KRA>
 A:Cross-references: GB:M29366
 R:Ploman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todar:
 Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
 A:Title: Molecular cloning and expression of another epidermal growth factor recept
 A:Reference number: 159164; PMID:90311312; PMID:2164210
 A:Accession: 159164
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-559; 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
 A:Cross-references: GB:M34309; NID:q183990; PIDN:AAA5979.1; PID:9306841
 C:Gene: GDB:ERBB3; HER3
 A:Cross-references: GDB:119880; OMIM:190151
 A:Map position: 12q13-12q13
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase t
 C:Keywords: ATP; phosphotransferase
 F:707-972/Domain: protein kinase homology <Kin>
 F:715-723/Region: protein kinase ATP-binding motif
 Query Match 35.0%; Score 2389.5; DB 2; Length 1342;
 Best Local Similarity 40.0%; Pred. No. 3.8e-91;
 Matches 526; Conservative 192; Mismatches 459; Indels 139; Gaps 33;
 QY 10 GLLALLPFGAA--STQVCTGTDKRLRLASEPETHDMLRLHYQCGQVQVQNLTYLPT 67
 Db 11 GLLFSIARGSEVNSQAVCPGLTNG:SVTGAENQYQTYLKYERREYVMGNLEIYLTGH 70
 QY 68 NASLSPLODIOEVQGVV:FAHQVQVPLQRLRIVRGTLQFEDNYVALAVLDGDDPLNNTT 127
 Db 71 NADLSLQWIREVTGYLVANNEFSLPLPNLRVVRGTQVYQKFAIFV-----LNNVT 125
 QY 128 PYTGASPGGLRELQRLSLTELKGGVLIQRNPOLCYODTILMKDIFHKNNQALTLIDTN 187
 Db 126 ----NSSHALROLRLTQLTEILISGVYIEKNKLCGMHDIWRDIVRDRD---AEIVXD 178
 QY 188 ASRACPGSPMKGRKCSWSESDCSLRTYACAGC-ARCKGPLPTDCHEQCAAGCTG 246
 Db 179 NGRSPCPCHVEVCKG-RCKWPGSEDCQTLTKTICAQQNGCHCGPNNQCHCECAGGCSG 237
 QY 247 PKHSDCLACIHNHSGICELACPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306
 Db 238 PDDTCFACRHHNDGACVPRCPPLVNVNKLFTGLEPNNHTKYQVGVVASCNPNFV-V 296
 QY 307 DVGSGTIVPEPLHNOEYTAEDGTORCEKSKPCARVVCYGLQMOYIRANSKF--IGITTELE- 363
 Db 297 DQTSQVRACPDPKMEVD-KNGLKCEPGGLCPKACBEGTG-----SGSRFQTVDSNIDG 350
 QY 364 FAGCKKIFGSLAFLEESFQGDPAASNTAPLOPQLOVFETLEITGYLYISAMPDPLPS 423
 Db 351 FVNCIKLIGNDLFTLTGLNDGPMHKIPALDEKLVNFTVREITGYLNIQSNPMMKNS 410
 QY 424 VFQNTQVIRGRILHNGAYS-LTLQGLG:SWGLRSLRSLGSLALIHNTLCLFVHTVPW 482


```

Db      411 VESNLTIGKSLYNGFSLIMKLVNTSLGFRSLKEISAGRIYISANRQLCVHSLNW 470
Qy      463 DQFNPRIQALLHTA-NRPEDECVGEGLACHOLCARGCMKGRPPQCNQCFLRGCECV 541
Db      471 TKVLGGPTPEERLDIKHNRPRRCDVAGKVCPLSSGGCGWGGPGQCLISCRNYSRGVCV 530
Qy      542 EECRVLOGLPREVYVARHCLPCHPECOFQNGSVTCFGEADQCVACAHYKDPFCVACRP 601
Db      531 THCNFLNPEPRFAHAEACFSCHPECCMEGATNGSGSDTCAQCAHFRGPHCVSSCP 590
Qy      602 SGVFPDLSYMPIMKRPDEBACQPCPINCISCVLDLDCGPAAE-----QRASPLTGI 654
Db      591 HGVILG--AKGPIYKPDVQNECRPCHENCTQCGKGPBELQDCIGQTLVILIGKTHLTMLAT 648
Qy      655 FNNFTVSPFLRVPKVSASHLEKRPQOKIRKVTMRRLQETLEVERLTPSGAMPQAOQARI 714
Db      649 IAGLVYIFMM-----IGTFTLYKRGRIQNKRAMRYLGRGESIEPLDS-ERANKVLARI 703
Qy      745 LKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRENTSFRANKELIDEAYVM 774
Db      704 FKETELRKIKVLGSGVGFVHKGVWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAI 763
Qy      775 AGVSPVYSRLIGICTSTVOLVTOLMPYGLDHYRENRGRISGODLLNMCQIAKMS 834
Db      764 GSLDAHAIYRELGLCPGSSQLVLYPLGSLDHYRQHGALGQLLNMGVQIAKMY 823
Qy      835 YLEDEVLYHRDLAARNVLYKSPNHVKITDFGLARLLDDETEYHADGGKVPKIMMALESI 894
Db      824 YLEHGMVHRNLAAANVLKSPQVADFGVADLPDDQQLLSEAKTPIKMMALESI 883
Qy      895 LRRFTQSDVMSYGVYTWELMTFGAKPYDGI PAEIPDLLEKGRLLQPPICITIDVYMI 954
Db      884 HFGKTYHOSDVMSYGVYTWELMTFGAEPYAGIRLAEVDPLEKGRLLQPPICITIDVYV 943
Qy      955 MWKCMNIDSECPREELVSEFRMARDPORFVVIQONEDLGA---SPLDSTFYRLIED 1011
Db      944 MWKCMNIDENIRPTKELANETRNARDPRLYLKRS-GRGIAPGEHPIGLNKLE 1002
Qy      1012 DDMGDLVDAEYLYVQOQFCPPDPAPGAGCMVHHNRSSSTRSGGGLTLGLEP-SEEA 1070
Db      1003 VELPELIDLDLLEED-----NLATTLTSLALSTLVGLTLNRPGR 1043
Qy      1071 PSPPLAPSEGASDVFDGLGMAKGLQSLPTHD-PSPLQYSSDPVPP-----SE 1123
Db      1044 SGLSLSPSSGY-MPMNGNLGSCQESAVSSSEKCPRVSLH-----PMRGLASBS 1096
Qy      1124 TDGYVA-----PLTCSQPE---YVQPDVPRQPSRPGP----- 1156
Db      1097 SEGHTGSEAELOEKVMSCRSRSPRPRGDSAYHSQRHSLTPTVPLSPGLEEDV 1156
Qy      1157 ----LPAARPAATLEBAKTSP-GRNGV-----KDVAFGCAVENPEYLLTPGGGAAP 1205
Db      1157 NGVMPDTHLKGTSSREGTSLSSVGLSYLGTDEED-----EEEYVNNRRRRSP 1208
Qy      1206 QHPPAPSPAFDNLVYMD-----QDPPERGAPSTFGTPTAENPEYL 1249
Db      1209 -FHPRPSSLEELGYEYMDVGSDSLASLGSTQSCGLHPVPIPTAGTTPDDDYEM 1263

```

RESULT 10

```

JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein, HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: J04387
R:Helixer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 273-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: J04387; MUID:96096535; PMID:8522190
A:Accession: J04387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>

```

```

A:Cross-references: GB:U29339; NID:9915389; PID:9915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GGT for res
A:Comment: This protein is a functional heregulin receptor that transduce signals
C:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase )
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prot
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (T
Query Match 33.7%; Score 2302.5; DB 2; Length 1339;
Best Local Similarity 40.3%; Pred. No. 148-87;
Matches 518; Conservative 169; Mismatches 439; Indels 159; Gaps 34;
Qy      3 LAAICRWGLLALLPPGAA---STQVCTGDMKRLRPASEPETHLDMRLHYOGCCOVQGN 59
Db      7 LQVLC-----FLSLARGSEMNGSAVCPGTLNGLSTGADNQGTYOTLYKLYKCEVVMGN 62
Qy      60 LELTYLPTNASLSFLQDIDQVQGVYLIANQVQVPLQRLIRYRGTQFEDNVALAVLN 119
Db      63 LELVLTGHNDLSFLQWIREVTVYVAVANNEFVPLPRLNRYVRGTQVYDGKRAIFVM-- 120
Qy      120 GDPINNTPTPTGASPGGLRDLQSLTEILKGGVLIORNPOLCYOPTILMKDIFHKNNQL 179
Db      121 ----LNTVY-----NSSHARQLKFTOLTLEISGVYIEKNDLCHMDITMRDLYRVR-- 170
Qy      180 ALTLIDTRSRACHPCSPMKGSRGWSESEDCQSLTRIVCAQC-ARCKGLPTDCHE 228
Db      171 GAETVYQNGANPCPCHEVCKG-RCKGPGEDCQILTKICAPQCKGRCPGPNQCCHD 229
Qy      239 QCAAGCGPBGSCCLAHNHSIGIELHCPALVTNTDFFESMPNDEGYTGASCVTA 228
Db      230 ECAAGCGPDTCCFARRNDGACVPRPRLVYNKLTFQLEPNHTTYQGVGCVAS 289
Qy      299 CPVNYLSTDVGSCTLVCPRLNDEVTAEQTQRCCKSKPCARVCYGL--GMQYIKANSKF 356
Db      290 CPHNFV-VDTFCVYACRPDKMEVD-KHGLKMCBPCGGLCPKACEGTGSGSRQYTDSSN 347
Qy      357 IGITELFPACCKIKFGLSLALPESFGDPRASNTAPLOPELOQVFTLEITGLIYLSAMP 416
Db      348 ID-----GFVACTKILGLDELITGLNVDPNPKI PALDPELVNFRVRELTYGLNTQSWP 403
Qy      417 DSLPLDSVFNQLGVIRGRIIANGAVS-LTQGLGISWLGRLSRELASGLALIHNTHL 475
Db      404 PHMNFVFNLTITIGRSLYNNGFSLIMKLVNTSLGFRSLKEISAGRIYISANQQLC 463
Qy      476 FHTVPWDQLFRNHQALLHTA-NRPEDECVGEGLACHOLCARGHCMGPGPTQCVNCSQF 534
Db      464 YHSLNMTLRLLRGSEERLDIKYDRPLGECIAEGKVCPLSSGGCGGCPAPGQCLSCRNY 523
Qy      535 LRQGEVCEBQVLOGLPREVYVARHCLPCHPECOFQNGSVTCFGEADQCVACAHYKDP 534
Db      524 SRGCVCTHGNFLQGEPRFVHEAQCSCHPECLPNEGISTYNGSGSDACARCAHFRPDG 563
Qy      595 FCVACPSGVKRPDLSYMPIMKRPDEBACQPCPINCISCVLDLDCGPAAEORASPLTS- 653
Db      584 HCVNCSCHGILG-AKGPITYKPDQAQNECRPCHENCTQCGNGBELDCLGQ--AEVLMSK 639
Qy      654 --IFNNTVSPFLRVPKVSASHLEKRPQOKIRKVTMRRLQETLEVERLTPSGAMPQAO 711
Db      640 PHVIAVAVGLAVLMLLGSSFLYMRGRQNKRRAMRYLERGESIEPLDS-ERANKVL 698
Qy      712 MRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRENTSFRANKELIDEA 771
Db      699 ARIFKETELRKVKVLGSGVGFVHKGVWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHM 758
Qy      772 YVAGVSPVYSRLIGICTSTVOLVTOLMPYGLDHYRENRGRISGODLLNMCQIAK 831

```

Db 759 LANGSDHAHIVLLGLCPSSSLQVLTQVPLGSLLDHVKQRETLGPOLLNMGVQIAK 818
 QY 832 GMSLEEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMAL 891
 Db 819 GMYTLEESHVHRDLALRNVMKSPQVQVADFGVADLLPPDDKQLHSEAKPIKIMAL 878
 QY 892 ESTLRFRFTHQSDVNSVGYVWMLMTFGAKPYDGIPIARLIPDLLEKGERLPDPPICTIDV 951
 Db 879 ESTHFGKRYTHQSDVNSVGYVWMLMTFGAEPRVAGRLAELPDLLEKGERLAPQICIIDV 938
 QY 952 YIMVVKCMIDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGPASPIDSTFYRSLIED 1011
 Db 939 YIMVVKCMIDENIRFTFKELAEFTRMARDPPIVIVIKAS-GPSTP-PAAPSVLTT 995
 QY 1012 DDMGDLVDAEYVLPQGFPCDPAPAGGMVHHRSSSTSGGDLTLGLEPSEE- 1068
 Db 996 KEL-GEAELEFEL- DLDDLEAELEGIA 1021
 QY 1069 -EAPRSPLAPSEG-AGSDVFDGDLGMGAAGKGLSLPT 1103
 Db 1022 TSLGSLSLPTGTLTPRSGQLSLSPSSGIMPMNOSLDBACLDASVLCGRQFSRPISL 1081
 QY 1104 HDSPLOKRYSEDTVLPSETDGV-APL-TC-SPOPE-YVNP 1143
 Db 1082 H-PIPRGR-PASESSEGHVTSSEALQEKVSVCRSRSRSPRRDGSAYHSQR 1133
 QY 1144 DVAPQPPSPREG-LEAPAPAGTILERAATLSP-GKNGV- 1185
 Db 1134 HSLTLVTPPLSPGLEEDNGVWMDTLRGASRSRBOGLSSVGLSGTEEBED- 1191
 QY 1186 AFGAVENPEYLTPOGGAAPQHP 1210
 Db 1192 -EYEMNKRGRSP-PRP 1209

RESULT 11

TFEVLV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C:Species: avian leukosis virus, ALV

C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999

C:Accession: B00643; A00643

C:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Citterenden, L.B.; Raines, M.

Cell 41, 719-726, 1985

A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro

A:Reference number: A00643; MUID:85228222; PMID:2988784

A:Accession: B00643

A:Molecule type: mRNA

A:Residues: 1-658 <NLT>

A:Cross-references: GB:M10066; GB:M13881; NID:G211749; PID:AAA48763.1; PID:G211750

A:Note: in Genbank entry CHKEB8B, release 109.0, the source is designated as Gallus gal

C:Comment: This protein is synthesized as a gag-env-erbB protein.

C:Genetics:

A:Gene: gag-env-erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific F

F.1-6/Product: gag protein (fragment) #status predicted <GAG>

F.7-59/Product: env protein (fragment) #status predicted <ENV>

F.60-69/Product: protein-tyrosine kinase erbB #status predicted <ERS>

F.194-459/Domain: protein kinase homology <KIN>

F.202-210/Region: protein kinase ATP-binding motif

F.229/Active site: Lys #status predicted

Query Match 25.2%; Score 1721.5; DB 1; Length 698;

Best Local Similarity 51.5%; Pred. No. 5,6e-64;

Matches 368; Conservative 73; Mismatches 151; Indels 123; Gaps 16;

QY 578 GEADQCVACAHYKPPFVACPSGVKPDLSYMIKFPPEDEGACQCPINCHSYVDL 637
 Db 60 GP-DHCKCAHFDGPHCVKACPAVGVGENDTL-VWXYADANAVCQCHPCTGCKGP 116
 QY 638 DKGCPAEQRASPLTSIFNNFTVSWFLRVPKVASHLKRRROKTRKTMRLLOETELV 697

Db 117 GLEGP--NSKTPSIAGVVGGLCLVVGIGLYLRRRHIVRKETLRRLLOERELV 173
 QY 698 EPLTPSGAMPQAOGRILKETELRKVKYLGSGAGFTYKGMIPDGEVVKIPVIMKYRE 757
 Db 174 EPLTPSGAMPQAOGRILKETELRKVKYLGSGAGFTYKGMIPDGEVVKIPVIMKYRE 233
 QY 758 NTSPPRANKEIDEAVYVAGVSPVSRLLGLTSTVOLVTLMPYGLLDHVENRGL 817
 Db 234 ATPSRANKEIDEAVYVAGVSPVSRLLGLTSTVOLVTLMPYGLLDHVENRGL 293
 QY 818 GSODLMMQMIAGMSVLEDRVLRDLAARNLVKSPNHVKITDFGLARLLDIDETEV 877
 Db 294 GSQYLLNMCVQIAGMNYLERRVLRDLAARNLVKTPQRYKITDFGLARLLDIDETEV 353
 QY 878 HADGKVPKIMMALESILRRFRFTHQSDVNSVGYVWMLMTFGAKPYDGIPIAREIPDLLEK 937
 Db 354 HADGKVPKIMMALESILRRFRFTHQSDVNSVGYVWMLMTGSKPYDGIPIAREIPDLLEK 413
 QY 938 GERLPQPICTIDVYIMVVKCMIDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGP 996
 Db 414 GERLPQPICTIDVYIMVVKCMIDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGP 473
 QY 997 ASPLDSTFYRSLIEDDDMGDLVDAEYVLPQGFPCDPAPAGGMVHHRSSSTSGG 1056
 Db 474 PSPIDSKRYRTLMEEDMEDIVDAEYVLPQGF- 513
 QY 1057 GDLTLGLEPSEERAPRSP-APSEGASDVFDGDLGMGAAGKGLSLPTHPSPLOK 1111
 Db 514 -SRTLLSLSTATNSNATNID- RNQGHVRRDSFQR 551
 QY 1112 YSEDPVLPSET-DGVAPLTCSPQPEYVNOVDVPRQPPSPREGPLPAAPAGATLER 1169
 Db 552 YSDPTGNFLEESIDDFGL-PAPEVNO-LMPKPS 585
 QY 1170 AKTLPCKNGVYXDF-AFGAVENPEYLTPOGGAAPQHP 1214
 Db 586 -TAMQONQIYNNISLTAISKLPMDSRVQNSHSTADNPEYL-NTQSLAK 634
 QY 1215 PAFNLYYWDQ-DPE- -RGAPSPFTFKGTPPTANEPEYGLDVP 1254
 Db 635 TVESSPYWIOGHNQINLDNPDYQODFLPRTKPNGLKVPALNPEYLRVAP 689

RESULT 12

TYUHL

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain)

C:Species: avian erythroblastosis virus

C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999

C:Accession: A00644; A38022

C:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.

Cell 35, 71-76, 1983

A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene

A:Reference number: A00644; MUID:84026539; PMID:61313229

A:Accession: A00644

A:Molecule type: DNA

A:Residues: 1-604 <YAM>

A:Cross-references: GB:M01216; NID:G209676; PID:AAA4400.1; PID:G209678

A:Note: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type

A:Reference number: A38022; MUID:84223957; PMID:61313229

A:Accession: A38022

A:Molecule type: DNA

A:Residues: 1-28, 'W', '30-139', 'F', '141-145', 'V', '147-152' <DEB>

A:Cross-references: GB:K02006

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific

F.118-135/Domain: protein kinase homology <KIN>

F.118-146/Region: protein kinase ATP-binding motif

F.165/Active site: Lys #status predicted

Query Match 24.3%; Score 1658; DB 1; Length 604;
 Best Local Similarity 51.4%; Pred. No. 2e-61;
 Matches 354; Conservative 69; Mismatches 142; Indels 124; Gaps 14;

587 CAHYKDPFCVAPCPSPKPDLSYMPKFPDEGACQPCPINCSTHSCVDLDDKGPAPQ 646
 3 CAHFDGPHCVKACPAVLGENDTL-VRKYADANAVALCHPCTROCKCPFGEGCP--- 58

647 RASPLTIFNNFTVSFRLRPKYSASHLEKRRQOKIRKYMRELLOTEVEPLTSGAM 706
 59 NGSTPSIAAGVGGI:CLVAVGIGLILYRRRIYKRLRLLORELEVEPLTSGEA 118

707 PNOQMRILKETELKRYKVLGSAFGTVYKGIWPDGENYKIPVAIKVLENTSPANK 766
 119 PNOHAKLKETEKRYKVLGSAFGTVYKGIWPDGENYKIPVAIKVLENTSPANK 178

767 ILDEAYVAGVGSFYSRLIGICITSTVQLTQMPRGCLLDHVRKRGSLGQDILNMC 826
 179 ILDEAYVAGVGSFYSRLIGICITSTVQLTQMPRGCLLDHVRKRGSLGQDILNMC 238

827 MQAKGKSYLEDVRLVARDLAARVAVKSPHVKITPFGALRLIDETEHADGKVP 886
 239 VOAKGMNLYEERLVARDLAARVAVKSPHVKITPFGALRLIDETEHADGKVP 298

887 KMALESILRRFTTHQSDWSYGVTVWELMTFGAKPYDGI:PAEIPDLKEXERLP 946
 299 KMALESILRRFTTHQSDWSYGVTVWELMTFGAKPYDGI:PAEIPDLKEXERLP 358

947 CTIDVYMTKCMWIDECRPRELVEFSRMRDQRFVIO-NEEDGAPAPLSTEF 1005
 359 CTIDVYMTKCMWIDECRPRELVEFSRMRDQRFVIO-NEEDGAPAPLSTEF 418

1006 RSLLEDDMDGLVDAEELVYPOQGFPCPDPAFGAGVHHRHSSSTRSGQDILGLEP 1065
 419 RSLLEDDMDGLVDAEELVYPOQGFPCPDPAFGAGVHHRHSSSTRSGQDILGLEP 449

1066 SEEAAPRSP-ASSEAGSDVFPDGLGMAKLOSLPTHDEPPLRYSEDPVPL 1120
 450 SEEAAPRSP-ASSEAGSDVFPDGLGMAKLOSLPTHDEPPLRYSEDPVPL 496

1121 PSEF--DGYVAPLTCSPQPEYVNPDPVPPPPPREPPLPAAPRAGATLEBRATLSPGN 1178
 497 LEESIDGFL-PAPEYVNO-LMPKKPSTAM----- 524

1179 GVYKDVAF-----GGAVENTEYLTOGGAAPRPPAPSPAFDN 1219
 525 GVYKDVAF-----GGAVENTEYLTOGGAAPRPPAPSPAFDN 575

1220 LYWDODPPERGAPSTFKGTPTAENPEY 1248
 576 LYWDODPPERGAPSTFKGTPTAENPEY 594

RESULT 13
 GQFFE
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbb
 C:Species: Drosophila melanogaster
 C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
 C:Accession: A00640; A38021
 R:Linne, B.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B. Z.
 Cell 40, 599-607, 1985
 A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
 A:Reference number: A00640; MUID:85124611; PMID:2982499
 A:Accession: A00640
 A:Molecule type: DNA
 A:Reids: 1-1330 <Lit>
 R:Madsworth, S. C.; Vincent III, W. S.; Bliedean, Wentworth, D.
 Nature 314, 178-180, 1985
 A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
 A:Reference number: A38021; MUID:85137938; PMID:2983232
 A:Accession: A38021

A:Molecule type: DNA
 A:Reids: 'A', 832-866 'V', 868-943, 'ORPSLVK' <WAD>
 A:Cross-references: EMBL:X02293; NID:G7922; PIDD:CA26157.1; PIDD:G929565
 C:Comment: This sequence is tentative because the introns have not been identified.
 C:Genetics:
 A:Gene: EGF
 A:Cross-references: FLYBase:FBgn0003731
 A:Map position: 2.57P
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph
 F:1-732/Domain: extracellular #status predicted <EX>
 F:733-764/Domain: transmembrane #status predicted <TM>
 F:765-1330/Domain: intracellular #status predicted <INT>
 F:808-1072/Domain: protein kinase homology <KIN>
 F:816-824/Region: protein kinase ATP-binding motif
 F:122-300,324,353,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic
 F:843/Active site: Lys #status predicted
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

Query Match 23.6%; Score 1611.5; DB 1; Length 1330;
 Best Local Similarity 29.6%; Pred. No. 3.5e-59;
 Matches 411; Conservative 175; Mismatches 414; Indels 387; Gaps 40;

80 VQGYVLAHNOYRQVRLRIRVSTQLE-EDNALAVLDNGDPLNTPVATGASP 134
 38 INYIVTGLDIPCTLSYRLQIRGRTLFSLVEEKYALFV-----TY 81

135 GGLRELQSLLEILKGVLTIOBNPOLCYQDTILMKDI:PHKNQOLATLIDNRSRACH 194
 82 SKMYTLEIPDLVDVNGVGFHNNYNLGMRTIQSEIVSNGTDAYVYDFAPRECK 141

195 CSPMKGSRKWESESSDCSLTRTVACAGCA--RCKGPLPTCCEQCAAGCTGPKHSDC 252
 142 CHESTCHG-CWGEGRKQCKSKLTCSPQACAGRCYGRKPRCHLFCAGCGTGTQKDC 200

253 LACLPHNSGICELHCPALVYNTDTFESMPNPEERYTPGASCUTACPPNYLSTVGSC 312
 201 IACKKFFDEAVSKCECPKPKRYNPTTYLETTEPEKRYVAGACVCEP-CHILRNGACV 259

313 LVCPHNOEVTADGTORCEKSKPCARVCYGLGQYIKANSKFIGTEL-----EPAG 366
 260 RSCPPQDKXDKGE-----CVPNGGCPKTC-----PGVYLAHGNIDSFN 300

367 CKKIFGSLAFLESFDG--DPAASNTA-----PIQPEQLQVEFLEITGYLYISAMPDS 418
 301 CTVIDGNIRILDQTFSGFODVYANTMGPRYIP:LPBERREVSITKEITGYLYINIEGTHPQ 360

419 LPDLSEFQNLQVIRIRILHNGAY-STLQGLISMLGLRSLELSGLALIHNNHLCFV 477
 361 PNLSTFRLLETTHGQLMESWPAALATVKSLSYSLERMKLQISSGVVIOHNDLCYV 420

478 HTVPMDQLFRNPQALLHTANRBEDEC----- 504
 421 SNIRWPAIQKEPEQKVMVNNLRADLCGLFLLISVQHNIMH:FAICREKMNHLGVS 480

505 ----- 504
 481 QGRLLGSHGVSVPYLQELQFQMLHRLMYLQVINSITDQKSNHQLTACYSAPVPT 540
 505 -----VG 506

541 SLITERARVAISAGLAMELEITARSASMRSKTLPAERGVPPWVFLGVCASARAGIA 600
 507 EGLA-----CHQLCARHCKGPGPTQCVNCSQFLRGQECVEECVLTQGLPREVY---N 556

601 EPLAGRAVCRKCHPCECTNNGYHEQVCSKCTHYKREDCETEC-----PADYTD 654
 557 ARHCLPCHPECCPQNGSVTCRGEADQVACAHYK-----DPRP-----CVARCSG 603

655 QRETCORHPEC---NG---CTGPADDCSKCNFLPLPAMETGPTVNSTMENCISKPLE 708
 604 VK-PDLSTYMPKFPDEGACQPCPINCSTHSCVDLDDKGPAPRASEPLTSIFNNFTVSF 662

```

Db      709 MRHVNQYTAIGY-----CAASPRSSKITANLD-----VMMIFITG 747
Qy      663 WLNVPKVASHLEK--RROOKIRKTY--MRLLQETELVEPLTPSGAMPNQOMILE 717
Db      748 AVLVPTICLCVTTYICRQOKAKETVKTMTALSGREDEPLRPINIGANLCKLIVD 807
Qy      718 TELRKVKVLSGAFGVYKGIWIPDGENVKI PVAIKVLRENTSFKANKIILDEAVYAGV 777
Db      808 AELRKGGVLSGMAFGVYKGVWVEGENVKI PVAIKELLKSTGAESEELREAYIMASE 867
Qy      778 GSPVSRILGICLTSTVOLVLTOLMPYGCILDHYENRGRGLSGODLLNMCKQAKGKSYLE 837
Db      868 EHVNLKLVAVCSQOMLITQLMPUGCLLDYVKNRDKISGALLNMSTQLAKGKSYLE 927
Qy      838 DVLVHRDLAARNVLYK---SPNHVKITDGLARLLDIDETEHADGKVPKIMWLES1 894
Db      928 EKLVHRDLAARNVLYRLLAGEDH---DFGLAKLLSSDSNEVKAKGKMPKIMWLECI 983
Qy      895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKERTLPORPCTIDVYMI 954
Db      984 RNEVFTSKSDVMAFGVYKGIWELTFGQRPHEINIPAKDI PDLIEVGLKEQPEICSDIYCT 1043
Qy      955 MVKCMIDSECRPRFRELVEFSRMARDPQRFVVIQNEDLG--PASPLDSTFYRSLLED 1012
Db      1044 LLSCHMLDAMRRTFFQLTTFVFAFARDPGRYALIGDKFTPLPA-----YTSODEKD 1096
Qy      1013 ---DWGDLVDAEELVLPQGFPCPPDAPGAGVHHRRSSSTRSGGDLTLGLEPSEEE 1069
Db      1097 LIRKLAPTTDGSEAIKRPDYLOPKALGPS-----HRTDCT-----DE 1135
Qy      1070 AP-----HSPLASGAGSDVFDG---DLGMAKGLQSLPTHDSPLQKRSSEPTVVL 1120
Db      1136 MPKLNRYCKDPKSNKNSGTGDDERDSSAREVGVNLR-----LDL 1174
Qy      1121 PSETDGYVALPTCSPOPEYVNOVDVRRPQPSREGLPARPAGATLERAKTSLPGKNGV 1180
Db      1175 PVDEDDYLP--TQCPGNNNNNNK-----NPNQNNMAVGAAGYM----- 1214
Qy      1181 VMDVFAFGAVENPEYL---TPQGAAPQPH-----PPAPSP-AP 1217
Db      1215 --DLIGPVSVNDPEYLLNAQTGVGESPIPTOTIGIPVMGPGTWEVKVPMGSEPTSS 1272
Qy      1218 DNLXYWD 1224
Db      1273 DHEYYND 1279

```

RESULT 14

S35745
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
 C:Species: avian erythroblastosis virus
 C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C:Accession: S35745
 R:Vennstrom, B.
 submitted to the EMBL Data Library, March 1993
 A:Reference number: S35743
 A:Accession: S35745
 A:Molecule type: DNA
 A:Residues: 1-544 <VEN>
 A:Cross-references: EMBL:X12707
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific F
 F:135-400/Domain: protein kinase homology <KIN>
 F:143-151/Region: protein kinase ATP-binding motif
 F:170/Active site: Lys #scatous predicted

Query Match 23.4%; Score 1602; DB 2; Length 544;
 Best Local Similarity 54.1%; Pred. No. 3,6e-59;
 Matches 339; Conservative 63; Mismatches 135; Indels 90; Gaps 13;

```

Qy      578 GPEADQVACAHYKDPPECVACRSGVYKPDLSYPMWKFPEDEGACQPCPINCTHSCVDL 637
Db      1 GP--DHCKCAHFIDGPHCVKACPAAGVGENDTL-VKXADANAVCQLCHPNCCTGCKGP 57
Qy      638 DDKGPAPQKASPLTSTINNTFTSWLRVVPKVASHLEKQOKIRKTYMRLLDETELV 697
Db      58 GLEGP---NSKTPSIAAGVVGGLCLVVGGLGIGLVRHHIRKTKTLLRLLQERELV 114
Qy      698 EPLTPSGAMPNQAKRILKETELRPVKVYLSGAFGVYKGIWIPDGENVKI PVAIKVLR 757
Db      115 EPLTPSGAMPNANHRLIKETEFKVKVYLSGAFGVYKGIWIPDGENVKI PVAIKVLR 174
Qy      758 NTPSPANKIILDEAVYAGVSPVYSRLLGICLTSTVOLVLTOLMPYGCILDHYENRGR 817
Db      175 ATSPFANKIILDEAVYAGVSPVYSRLLGICLTSTVOLVLTOLMPYGCILDHYENRGR 234
Qy      818 GSODLLNMCKQAKGKSYLEVDRLVHNPRLAARNVLYKSPNHVKITDGLARLLDIDETEH 877
Db      235 GSQYLLNMCKQAKGKSYLEVDRLVHNPRLAARNVLYKSPNHVKITDGLARLLDIDETEH 294
Qy      878 HADGKVPKIMWLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEK 937
Db      295 HADGKVPKIMWLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEK 354
Qy      938 GBRLPORPCTIDVYMIWYKCMIDSECRPRFRELVEFSRMARDPQRFVVIQ-NEDIGP 996
Db      355 GERLPORPCTIDVYMIWYKCMIDSECRPRFRELVEFSRMARDPQRFVVIQ-NEDIGP 414
Qy      997 ASPLDSTFYRSLLEDMDGDLVDAEELVLPQGFPCPPDAPGAGVHHRRSSSTRSGG 1056
Db      415 PSFTDSKPYRLMEEDMEDIVDADEVLPQGF-----NSPST----- 454
Qy      1057 GDLTLGLEPSEEAERSPPL-----APSEGAGSDVFDGDLGMAKGLQSLPTHDSPLQ 1111
Db      455 -----SRTPLLSISATSNKSNATNCIDRNG-----H----- 481
Qy      1112 YSEDPVTLPSBETDGYVALPTCSPOPEYVNOVDVRRPQPSREGLPARPAGAT-LEPA 1170
Db      482 -----PYREOGFL-----PAPEYVNO--LMPKPSSTAMVNOIYNYISLTAKSL 524
C      1171 KTLSPKNGVYKDVFAFGAVENPEYL 1197
Db      525 PIDSRYN-----SHSTAVDNPEYL 544

```

RESULT 15

S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
 C:Species: avian erythroblastosis virus
 C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C:Accession: S00727
 R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus m
 A:Reference number: S00727; MUID:88217326; PMID:2897102
 A:Accession: S00727
 A:Molecule type: DNA
 A:Residues: 1-545 <SCO>
 A:Cross-references: EMBL:X06943
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; phosphotransferase
 F:135-400/Domain: protein kinase homology <KIN>
 F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.3%; Score 1595; DB 2; Length 545;
 Best Local Similarity 54.1%; Pred. No. 7e-59;
 Matches 339; Conservative 62; Mismatches 136; Indels 90; Gaps 13;
 Qy 578 GPEADQVACAHYKDPPECVACRSGVYKPDLSYPMWKFPEDEGACQPCPINCTHSCVDL 637
 Db 1 GP--DHCKCAHFIDGPHCVKACPAAGVGENDTL-VKXADANAVCQLCHPNCCTGCKGP 57

```
QY 638 DDGCPAEQASPLTIFNNFTVSFWLVRPKVSASHLEKRRQCKIRKRYTMRRLQETELY 697
   |||
   |||
Db 58 GLEGC---NGSKTPSIAAGVWGGLCLVVGGLGIGLYLRRIHVRKRIIRLQRELY 114
   |||
   |||
QY 698 EPLTPSGAMPNQAMRIKETEIRKVKVLGSGAFGYKGIWIPDGENVKIPVAIKVLR 757
   |||
   |||
Db 115 EPLTPSGEAPNQAHRIKETEIRKVKVLGSGAFGYKGIWIPDGENVKIPVAIKVLR 174
   |||
   |||
QY 758 NTPSPKANKELDEAYVAGVSPVSRLLGICLTSTQVLTOLMPYGCILDHRENRGL 817
   |||
   |||
Db 175 ATSPKANKETLDEAYVAGVSPVSRLLGICLTSTQVLTOLMPYGCILDHRENRGL 234
   |||
   |||
QY 818 GSODLLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEX 877
   |||
   |||
Db 235 GSQYLLNMCQIAKMSYLEDVRLVHRDLAARNVLYKTPQDVKITDGLAKOLGADEXEY 294
   |||
   |||
QY 878 HADGKVPKIMMALESILRRRFTHQSDVMSGYTVWELMTFGAKPYDGIIPAREIPDLLEK 937
   |||
   |||
Db 295 HAEGKVPKIMMALESILRRRFTHQSDVMSGYTVWELMTFGSKPYDGIIPAREIPDLLEK 354
   |||
   |||
QY 938 GERLPQPPICITIDVYIMVQWIMIDSECRPRFRELVSFSGRMARDPQRFVYIQ-NEDLGP 996
   |||
   |||
Db 355 GERLPQPPICITIDVYIMVQWIMIDSECRPRFRELVSFSGRMARDPQRFVYIQ-NEDLGP 414
   |||
   |||
QY 997 ASPLDSTFYRLLEDMDMDLVDAEELVLPQGFPCPDPAAGAGMTHRRHRSSTRSGG 1056
   |||
   |||
Db 415 PSPTDSKFRYRLMEEDMEDIVDAEYLVPHQGF-----NSPST---- 454
   |||
   |||
QY 1057 GDULGLEPSEEEAPRSPL-----APSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOK 1111
   |||
   |||
Db 455 -----SRTPLLSLSLATSNSATNCIDRNG-----H----- 481
   |||
   |||
QY 1112 YSEDPVPLPSETDGYVAPLTCSPQPEYVNPQPPRQPSREGPLPAARPAAT-LERA 1170
   |||
   |||
Db 482 -----PVREDGFL-----PAPEYVNPQ--LMPKPPSTAVVQVQVYISLTAISKU 524
   |||
   |||
QY 1171 KTLSPKNGVYKDYFAGGAVENPEYL 1197
   |||
   |||
Db 525 PMDSRYQN-----SHSTAVDNPEYL 544
   |||
   |||
```

Search completed: July 22, 2003, 09:27:28
Job time : 30.2855 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 Seconds

(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-632-652-14

Sequence: 1 MELALCRMGJLIALPPCA.....TFKPTPTAENPEYLGLDPEV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6623	96.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5865	85.8	1257	1 ERB2_RAT	P06494 rattus norv
3	5841.5	85.5	1254	1 ERB2_MESAU	P06053 mesocricetu
4	3111	45.5	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3081	45.1	1210	1 EGFR_MOUSE	P01379 mus musculu
6	2920.5	42.7	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2906	42.5	1308	1 ERB4_RAT	Q62356 rattus norv
8	2643.5	38.7	1167	1 XMRK_XIPMA	P13388 xiphiophorus
9	2398.5	35.1	1342	1 ERB3_HUMAN	P21660 homo sapien
10	2325.5	34.0	1339	1 ERB3_RAT	Q62799 rattus norv
11	1936	28.3	1426	1 EGFR_DROME	P04412 drosophila
12	1704.5	24.9	634	1 ERBB_ALV	P00534 avian leuko
13	1658	23.3	604	1 ERBB_AVIER	P00535 avian eryth
14	1585	22.7	540	1 ERBB_AVIEU	P11373 avian eryth
15	1550	22.7	703	1 EGFR_CHICK	P13387 gallus gal1
16	1274	18.6	1333	1 LTR3_CAEL	P24348 caenorhabdi
17	1142.5	16.7	245	1 ERB2_MOUSE	P70324 mus musculu
18	713	10.4	1363	1 ILPR_BRALA	O02466 brachiolesto
19	693.5	10.1	1372	1 INSR_MOUSE	P15208 mus musculu
20	692	10.1	1300	1 INSR_MOUSE	Q94514 mus musculu
21	691	10.1	1382	1 INSR_HUMAN	P06127 rattus norv
22	680	10.1	1383	1 INSR_RAT	P14616 homo sapien
23	687	10.1	1297	1 IRR_HUMAN	P14617 cavia porce
24	680.5	10.0	1300	1 IRR_CAVPO	Q25197 hydra atten
25	680	10.0	1477	1 HTR7_HYDAT	O25410 lymnaea sca
26	680	10.0	1607	1 MPR_LYMAT	P08069 homo sapien
27	644	9.4	1367	1 IGIR_HUMAN	Q60751 mus musculu
28	630	9.2	1373	1 IGIR_MOUSE	P24062 rattus norv
29	626.5	9.2	1370	1 IGIR_RAT	Q93105 aedes aegypt
30	617	9.0	1390	1 INSR_AEDAE	P09208 drosophila
31	616	8.6	1114	1 INSR_DROME	P07949 homo sapien
32	591	8.6	1053	1 RET_HUMAN	O00344 gallus gall
33	589.5	8.6	1053	1 FAK1_CHICK	

34	582	8.5	757	1 HTR6_HYDAT	P53356 hydra atten
35	579	8.5	1052	1 FAK1_MOUSE	P34152 mus musculu
36	578.5	8.5	984	1 EPB1_CHICK	O07494 gallus gall
37	576.5	8.4	984	1 EPB1_RAT	P09759 rattus norv
38	576	8.4	1055	1 FAK1_RAT	O03346 rattus norv
39	572	8.4	1052	1 FAK1_HUMAN	O05397 homo sapien
40	570.5	8.3	984	1 EPB1_HUMAN	P54762 homo sapien
41	570	8.3	987	1 EPB4_HUMAN	P54760 homo sapien
42	569	8.3	902	1 EPBB_XENLA	Q91736 xenopus lae
43	569	8.3	1068	1 FAK1_XENLA	O91738 xenopus lae
44	555.5	8.1	1009	1 FAK2_RAT	P70600 rattus norv
45	554.5	8.1	746	1 ABL_MLVAB	P00521 abelson mur

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD: PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (P15setB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.,
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234 (1986).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.,
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139 (1985).
RN (3)
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=295967;
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RL human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).
RN (4)
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429 (1993).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; M11767; AAA35808.1; -.
 DR EMBL; M11761; AAA35808.1; JOINED.
 DR EMBL; M11762; AAA35808.1; JOINED.
 DR EMBL; M11763; AAA35808.1; JOINED.
 DR EMBL; M11764; AAA35808.1; JOINED.
 DR EMBL; M11765; AAA35808.1; JOINED.
 DR EMBL; M11766; AAA35808.1; JOINED.
 DR EMBL; M11730; AAA35808.1; -.
 DR EMBL; M12036; AAA35808.1; -.
 DR EMBL; X03563; CAA27060.1; -.
 DR PIR; A25491; A25491.
 DR PIR; A24571; A24571.
 DR HSSP; P11362; 1FCG.
 DR Genem; HGNC:3430; ERBB2.
 DR MIM; 164870; -.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene Family; Receptor; Signal;
 DR Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 DR Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 720 987 PROTEIN KINASE.
 FT BINDING 726 734 ATP (BY SIMILARITY).
 FT ACT_SITE 753 753 ATP (BY SIMILARITY).
 FT DISULFID 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	634	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTID=VAR_004077.
FT	VARIANT	655	655	I -> V.
FT	CONFLICT	1170	1170	/FTID=VAR_004078.
FT	CONFLICT	1170	1170	P -> A (IN REF. 2).
SEQ	SEQUENCE	1255 AA; 137909 MW; 39B9DFDA04C9F62 CRC64;		

Query Match 96.9%; Score 6623; DB 1; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0; Mismatches 28; Indels 0; Gaps 0;
 Matches 1221; Conservative 6;

QY	1	MELALCRMGLLALPFGAASVCTGTGDMRLPASPETHLDMRLHLYGCGVQGNL	60
QY <td>1</td> <td>MELALCRMGLLALPFGAASVCTGTGDMRLPASPETHLDMRLHLYGCGVQGNL</td> <td>60</td>	1	MELALCRMGLLALPFGAASVCTGTGDMRLPASPETHLDMRLHLYGCGVQGNL	60
QY <td>61</td> <td>ELTYPTNASTSFLQDIQGVGVLIANQVQVLRIRYRGVQLFEDNYALAVLNG</td> <td>120</td>	61	ELTYPTNASTSFLQDIQGVGVLIANQVQVLRIRYRGVQLFEDNYALAVLNG	120
QY <td>61</td> <td>ELTYPTNASTSFLQDIQGVGVLIANQVQVLRIRYRGVQLFEDNYALAVLNG</td> <td>120</td>	61	ELTYPTNASTSFLQDIQGVGVLIANQVQVLRIRYRGVQLFEDNYALAVLNG	120
QY <td>121</td> <td>DELNNTPTVTGASPGGLRELQRSLEILKGVLLQRPOLQYQDTILMKDIFHNQOLA</td> <td>180</td>	121	DELNNTPTVTGASPGGLRELQRSLEILKGVLLQRPOLQYQDTILMKDIFHNQOLA	180
QY <td>121</td> <td>DELNNTPTVTGASPGGLRELQRSLEILKGVLLQRPOLQYQDTILMKDIFHNQOLA</td> <td>180</td>	121	DELNNTPTVTGASPGGLRELQRSLEILKGVLLQRPOLQYQDTILMKDIFHNQOLA	180
QY <td>121</td> <td>DELNNTPTVTGASPGGLRELQRSLEILKGVLLQRPOLQYQDTILMKDIFHNQOLA</td> <td>180</td>	121	DELNNTPTVTGASPGGLRELQRSLEILKGVLLQRPOLQYQDTILMKDIFHNQOLA	180
QY <td>181</td> <td>LFLIDNRSRACHPCSPMKCSRCWGESESDCSLTRVCAGGACARCKGPLETDCHEQC</td> <td>240</td>	181	LFLIDNRSRACHPCSPMKCSRCWGESESDCSLTRVCAGGACARCKGPLETDCHEQC	240
QY <td>181</td> <td>LFLIDNRSRACHPCSPMKCSRCWGESESDCSLTRVCAGGACARCKGPLETDCHEQC</td> <td>240</td>	181	LFLIDNRSRACHPCSPMKCSRCWGESESDCSLTRVCAGGACARCKGPLETDCHEQC	240
QY <td>181</td> <td>LFLIDNRSRACHPCSPMKCSRCWGESESDCSLTRVCAGGACARCKGPLETDCHEQC</td> <td>240</td>	181	LFLIDNRSRACHPCSPMKCSRCWGESESDCSLTRVCAGGACARCKGPLETDCHEQC	240
QY <td>241</td> <td>AAGCTGPRKSDCLCLHPHSGICGLHCPALVTNTDFFESNPNBEGRTFASCVTACP</td> <td>300</td>	241	AAGCTGPRKSDCLCLHPHSGICGLHCPALVTNTDFFESNPNBEGRTFASCVTACP	300
QY <td>241</td> <td>AAGCTGPRKSDCLCLHPHSGICGLHCPALVTNTDFFESNPNBEGRTFASCVTACP</td> <td>300</td>	241	AAGCTGPRKSDCLCLHPHSGICGLHCPALVTNTDFFESNPNBEGRTFASCVTACP	300
QY <td>241</td> <td>AAGCTGPRKSDCLCLHPHSGICGLHCPALVTNTDFFESNPNBEGRTFASCVTACP</td> <td>300</td>	241	AAGCTGPRKSDCLCLHPHSGICGLHCPALVTNTDFFESNPNBEGRTFASCVTACP	300
QY <td>301</td> <td>VNVLSTDVSSCTLVCPPLHNOEYTAEDGTQRCESKPCARVCYGLGMQYIKANSKFIGIT</td> <td>360</td>	301	VNVLSTDVSSCTLVCPPLHNOEYTAEDGTQRCESKPCARVCYGLGMQYIKANSKFIGIT	360
QY <td>301</td> <td>VNVLSTDVSSCTLVCPPLHNOEYTAEDGTQRCESKPCARVCYGLGMQYIKANSKFIGIT</td> <td>360</td>	301	VNVLSTDVSSCTLVCPPLHNOEYTAEDGTQRCESKPCARVCYGLGMQYIKANSKFIGIT	360
QY <td>301</td> <td>VNVLSTDVSSCTLVCPPLHNOEYTAEDGTQRCESKPCARVCYGLGMQYIKANSKFIGIT</td> <td>360</td>	301	VNVLSTDVSSCTLVCPPLHNOEYTAEDGTQRCESKPCARVCYGLGMQYIKANSKFIGIT	360
QY <td>361</td> <td>ELIEFAGCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLEITGYLYISAMPDILP</td> <td>420</td>	361	ELIEFAGCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLEITGYLYISAMPDILP	420
QY <td>361</td> <td>ELIEFAGCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLEITGYLYISAMPDILP</td> <td>420</td>	361	ELIEFAGCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLEITGYLYISAMPDILP	420
QY <td>361</td> <td>ELIEFAGCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLEITGYLYISAMPDILP</td> <td>420</td>	361	ELIEFAGCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLEITGYLYISAMPDILP	420
QY <td>421</td> <td>DLSTVQNLQVINGRIHNGAVSLTLQGGISMLGRLSRELGSGALLHHNTHLQCFVTV</td> <td>480</td>	421	DLSTVQNLQVINGRIHNGAVSLTLQGGISMLGRLSRELGSGALLHHNTHLQCFVTV	480
QY <td>421</td> <td>DLSTVQNLQVINGRIHNGAVSLTLQGGISMLGRLSRELGSGALLHHNTHLQCFVTV</td> <td>480</td>	421	DLSTVQNLQVINGRIHNGAVSLTLQGGISMLGRLSRELGSGALLHHNTHLQCFVTV	480
QY <td>421</td> <td>DLSTVQNLQVINGRIHNGAVSLTLQGGISMLGRLSRELGSGALLHHNTHLQCFVTV</td> <td>480</td>	421	DLSTVQNLQVINGRIHNGAVSLTLQGGISMLGRLSRELGSGALLHHNTHLQCFVTV	480
QY <td>481</td> <td>PWDQLFRRNHQALLHTANRPEDECVGEGIAHQLCARHGWPGPTQCVNCSQFLRGEC</td> <td>540</td>	481	PWDQLFRRNHQALLHTANRPEDECVGEGIAHQLCARHGWPGPTQCVNCSQFLRGEC	540
QY <td>481</td> <td>PWDQLFRRNHQALLHTANRPEDECVGEGIAHQLCARHGWPGPTQCVNCSQFLRGEC</td> <td>540</td>	481	PWDQLFRRNHQALLHTANRPEDECVGEGIAHQLCARHGWPGPTQCVNCSQFLRGEC	540
QY <td>481</td> <td>PWDQLFRRNHQALLHTANRPEDECVGEGIAHQLCARHGWPGPTQCVNCSQFLRGEC</td> <td>540</td>	481	PWDQLFRRNHQALLHTANRPEDECVGEGIAHQLCARHGWPGPTQCVNCSQFLRGEC	540
QY <td>541</td> <td>VEECVQLQCLPREYNARHCLPCHPECPQNGSVTCFPEADQCAACHYDPPCVARC</td> <td>600</td>	541	VEECVQLQCLPREYNARHCLPCHPECPQNGSVTCFPEADQCAACHYDPPCVARC	600
QY <td>541</td> <td>VEECVQLQCLPREYNARHCLPCHPECPQNGSVTCFPEADQCAACHYDPPCVARC</td> <td>600</td>	541	VEECVQLQCLPREYNARHCLPCHPECPQNGSVTCFPEADQCAACHYDPPCVARC	600
QY <td>541</td> <td>VEECVQLQCLPREYNARHCLPCHPECPQNGSVTCFPEADQCAACHYDPPCVARC</td> <td>600</td>	541	VEECVQLQCLPREYNARHCLPCHPECPQNGSVTCFPEADQCAACHYDPPCVARC	600
QY <td>601</td> <td>PSGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAPQORASPLTSIFNNFTV</td> <td>660</td>	601	PSGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAPQORASPLTSIFNNFTV	660
QY <td>601</td> <td>PSGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAPQORASPLTSIFNNFTV</td> <td>660</td>	601	PSGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAPQORASPLTSIFNNFTV	660
QY <td>601</td> <td>PSGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAPQORASPLTSIFNNFTV</td> <td>660</td>	601	PSGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAPQORASPLTSIFNNFTV	660
QY <td>661</td> <td>SWLRVPKYSASHLEKROQKIRKTYMARLLQETLVEPLTPSGAMPQOQRIKTEL</td> <td>720</td>	661	SWLRVPKYSASHLEKROQKIRKTYMARLLQETLVEPLTPSGAMPQOQRIKTEL	720
QY <td>661</td> <td>SWLRVPKYSASHLEKROQKIRKTYMARLLQETLVEPLTPSGAMPQOQRIKTEL</td> <td>720</td>	661	SWLRVPKYSASHLEKROQKIRKTYMARLLQETLVEPLTPSGAMPQOQRIKTEL	720
QY <td>661</td> <td>SWLRVPKYSASHLEKROQKIRKTYMARLLQETLVEPLTPSGAMPQOQRIKTEL</td> <td>720</td>	661	SWLRVPKYSASHLEKROQKIRKTYMARLLQETLVEPLTPSGAMPQOQRIKTEL	720

QY	721	RKVKVLGSGAFSTYKGI	IPDGENVKI	PVALIKVIRENTP	SPANKNEI	IDEAYVMAGVSP	780				
Db	721	RKVKVLGSGAFSTYKGI	IPDGENVKI	PVALIKVIRENTP	SPANKNEI	IDEAYVMAGVSP	780				
QY	781	YYSRLIGTCLSTVOLV	OLMPYGC	LDHVRENRCL	SGODLNN	CMOIAKMSYLEDVR	840				
Db	781	YYSRLIGTCLSTVOLV	OLMPYGC	LDHVRENRCL	SGODLNN	CMOIAKMSYLEDVR	840				
QY	841	LVHRDLAARNVLSK	PNHVKITD	FGARLLD	IDETEVYADGKVP	IKMMALESLRRFT	900				
Db	841	LVHRDLAARNVLSK	PNHVKITD	FGARLLD	IDETEVYADGKVP	IKMMALESLRRFT	900				
QY	901	HOSDVMSCGVYWE	LMITFGAKPYDG	IPAREI	IDLEKEBERLP	POPICTIDVYMIWKCMM	960				
Db	901	HOSDVMSCGVYWE	LMITFGAKPYDG	IPAREI	IDLEKEBERLP	POPICTIDVYMIWKCMM	960				
QY	961	IDSECRPREPREV	SEFSRKAARD	PORFVY	IONEDLGPAPL	PLSTFPRSIL	LEDDMDGLDVA	1020			
Db	961	IDSECRPREPREV	SEFSRKAARD	PORFVY	IONEDLGPAPL	PLSTFPRSIL	LEDDMDGLDVA	1020			
QY	1021	BEYLVPOQGF	PCPDPA	PAGAGM	VHHRHSS	STRSGGDL	TLGLBSEBEARSP	PLAPSEG	1080		
Db	1021	BEYLVPOQGF	PCPDPA	PAGAGM	VHHRHSS	STRSGGDL	TLGLBSEBEARSP	PLAPSEG	1080		
QY	1081	AGSDVFDGDL	GMGAKKGI	OSLPT	HPSPLO	RSEPTPL	PLSETDGYA	PLTCSPOREYV	1140		
Db	1081	AGSDVFDGDL	GMGAKKGI	OSLPT	HPSPLO	RSEPTPL	PLSETDGYA	PLTCSPOREYV	1140		
QY	1141	NQPDVRPOP	PPSPRE	CPPLAAR	PAGATLE	RAKTLSP	GKNVYKDV	FAFGA	VENBEYLTPO	1200	
Db	1141	NQPDVRPOP	PPSPRE	CPPLAAR	PAGATLE	RAKTLSP	GKNVYKDV	FAFGA	VENBEYLTPO	1200	
QY	1201	GGAAPQHP	PPPAFSP	AFDNLY	WQDDP	PERGAP	PSSTFG	CTAEN	PEYLG	LDVRY	1255
Db	1201	GGAAPQHP	PPPAFSP	AFDNLY	WQDDP	PERGAP	PSSTFG	CTAEN	PEYLG	LDVRY	1255

	RA	Newman R Crampton M.J., Stenbery M.J.E., Campbell I.D. ; "Three dimensional structure of the transmembrane region of the proto-
	RT	oncogenic and oncogenic forms of the neu protein."
	RL	EMBO J. 11:43-48(1992).
	CC	-I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
	CC	ALTHOUGH NEREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
	CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
	CC	ALPHA AND ANGIOREGULIN.
	CC	-I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
	CC	tyrosine phosphate.
	CC	-I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
	CC	THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
	CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.
	CC	-I- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
	CC	RESIDUES (BY SIMILARITY).
	CC	-I- SIMILARITY: BELONGS TO THE EGFR RECEPTOR FAMILY. -----
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). -----
	CC	EMBL; X03362; CAA27059.1 / ALT_INIT.
	DR	PIR; A24562; TVRTNU.
	DR	HSSP; P11362; IFK.
	DR	InterPro; IPRO000494; EGFR_L_domain.
	DR	InterPro; IPRO00719; Euk_Pkinase.
	DR	InterPro; IPRO02174; Furin-like.
	DR	InterPro; IPRO01245; Tyr_pkinase.
	DR	InterPro; IPRO04019; YLP_motif.
	DR	pfam; PF00069; pkinase; 1.
	DR	pfam; PF00757; Furin-like; 1.
	DR	pfam; PF01030; Recep_L_domain; 2.
	DR	pfam; pf02757; ylp_2.
	DR	pfdom; PD000001; Euk_pkinase; 1.
	DR	SMART; SMO0261; FU; 3.
	DR	SMART; SMO0219; TYRK; 1.
	DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
	DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
	DR	PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
	KW	Transmembrane, Glycoprotein, Multigene family, Receptor, Signal; Transferase, Tyrosine-protein kinase, ATP-binding, Phosphorylation, Proto-oncogene, Disease mutation.
	KW	SIGNAL 1 21 POTENTIAL.
	FT	CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
	FT	DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
	FT	TRANSMEM 655 677 POTENTIAL.
	FT	CYS-RICH CYTOPLASMIC (POTENTIAL).
	FT	DOMAIN 678 1257 CYS-RICH.
	FT	DOMAIN 159 369 CYS-RICH.
	FT	DOMAIN 473 646 PROTEIN KINASE.
	FT	NP_BIND 722 989 ATP (BY SIMILARITY).
	FT	BINDING 728 736 ATP (BY SIMILARITY).
	FT	BINDING 755 755 BY SIMILARITY.
	FT	ACT_SITE 847 847 BY SIMILARITY.
	FT	DISEUFD 196 205 BY SIMILARITY.
	FT	DISEUFD 200 213 BY SIMILARITY.
	FT	DISEUFD 221 228 BY SIMILARITY.
	FT	DISEUFD 225 236 BY SIMILARITY.
	FT	DISEUFD 237 245 BY SIMILARITY.
	FT	DISEUFD 241 253 BY SIMILARITY.
	FT	DISEUFD 256 265 BY SIMILARITY.
	FT	DISEUFD 269 296 BY SIMILARITY.
	FT	DISEUFD 300 312 BY SIMILARITY.
	FT	DISEUFD 316 332 BY SIMILARITY.
	FT	DISEUFD 335 339 BY SIMILARITY.
	FT	DISEUFD 513 522 BY SIMILARITY.
	FT	DISEUFD 517 530 BY SIMILARITY.
	FT	DISEUFD 533 542 BY SIMILARITY.
	FT	DISEUFD 546 562 BY SIMILARITY.
	FT	DISEUFD 565 578 BY SIMILARITY.

```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 642 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.8%; Score 5665; DB 1; Length 1257;
Best Local Similarity 86.2%; Pred. No. 1,6e-302;
Matches 1084; Conservative 48; Mismatches 123; Indels 2; Gaps 2;

QY 1 MELALCWMGLLLALPPGAASVCTGTDMKRLPASPETHLDMRLYQGCQVVGQNL 60
DB 1 MELAAWCKGFLLLPPIAGTCTGTDMKRLPASPETHLDMRLYQGCQVVGQNL 60
QY 61 ELTYPTNASLFLQIDQEVGVLIANQVQVQLRLRTVGTQLFEDNVALVDNG 120
DB 61 ELTYVPANASLFLQIDQEVGVLIANQVQVQLRLRTVGTQLFEDNVALVDNR 120
QY 121 DELNNTPTVT-GASPGGLRELRLRSTELTKGVLIQNRNPOLCYODTLMKDIPKNNQL 179
DB 121 DQDNVAATPRTGREGLELDRLSTELTKGVLIQNRNPOLCYODTLMKDIPKNNQL 180
QY 180 ALTLIDNRSRACHPCSPCKGSRGWSSESDQSLRTVCAGGACRKGPLPTDCHEQ 239
DB 181 AFDVIDNRSRACPCAPACKKNHGWSESDQSLRTVCAGGACRKGPLPTDCHEQ 240
QY 240 CAAGCTGPHGSDCLCLHFNHSGICELHCPALVTNTPTFSEMPBERVYFGASCTAC 299
DB 241 CAAGCTGPHGSDCLCLHFNHSGICELHCPALVTNTPTFSEMPBERVYFGASCTTC 300
QY 300 PNYLSTDVSGCTLVCPHNOEVTAEADGTORCEKSKPCARCYGLQNYIKANSKFTGI 359
DB 301 PNYLSTEVSGCTLVCPHNOEVTAEADGTORCEKSKPCARCYGLQNYIKANSKFTGI 360
QY 360 TELERAGCKKIFGSLAFLESPDGPASTAPLOBOUOVFTLEITIGLYISAMPDSL 419
DB 361 NVQEPDGGCKKIFGSLAFLESPDGPASTAPLOBOUOVFTLEITIGLYISAMPDSL 420
QY 420 PDLVYFONLQVIRGRILHNGAYSLTQGLGISMGLRSIRELSGLALIHNTHLCPVHT 479
DB 421 RDLVYFONLRIIRGRILHNGAYSLTQGLGISMGLRSIRELSGLALIHNTHLCPVHT 480
QY 480 VWMDOLEFNFHQAALHTANRPDE-CVGBGLACHOLCARHGWGFGPTQVCNCSQFLRG 538
DB 481 VWMDOLEFNFHQAALHTANRPDE-CVGBGLACHOLCARHGWGFGPTQVCNCSHFLRG 540
QY 539 ECVESCRVLOGLPREVYVARHCLPCHPEQOPONGSVTGPBPADCCVCAHXPDPFVA 598
DB 541 ECVESCRVWKGIPREYVSDKCLPCHPEQOPONSETGCGSADCAAHYKSSSCVA 600
QY 599 RCPGSGVPLSLYMPIMKPPDEEGACQPCPINCTHSCVDLDKGPABRASPPLTSIPNNF 658
DB 601 RCPGSGVPLSLYMPIMKPPDEEGACQPCPINCTHSCVDLDKGPABRASPPLTSIPNNF 660
QY 659 TVSFVLVYKVSASHLEKRRKOKIRKYMRLLOSTELVEPLTSGAMPNOQMKILET 718
DB 661 VGVLLFLIVVVGILIKRRKOKIRKYMRLLOSTELVEPLTSGAMPNOQMKILET 720
QY 719 ELRKVKVLSGAGFVYKGIWIPDENYKIPVAIVLENTSPKANKIILDEAYVMAGVG 778
DB 721 ELRKVKVLSGAGFVYKGIWIPDENYKIPVAIVLENTSPKANKIILDEAYVMAGVG 780
QY 779 SFYVSRLLIGLICLTIVQLVTOIMPYGCILDHVENRGRIGSODLLMCMQIAKGMSTYED 838

```

```

DB 781 SFYVSRLLIGLICLTIVQLVTOIMPYGCILDHVENRGRIGSODLLMCMQIAKGMSTYED 840
QY 839 VRLVHRLARNLVNSPNNVKTTPDELALLIDIEFHYADGKVPKMALESILRR 898
DB 841 VRLVHRLARNLVNSPNNVKTTPDELALLIDIEFHYADGKVPKMALESILRR 900
QY 899 FTHOSDWSGVYTWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTIDVYMIWKC 958
DB 901 FTHOSDWSGVYTWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTIDVYMIWKC 960
QY 959 WMIDSECRPPFRELVSFSSMARDPQFVVIQNEIDGAPASPLDSTFYRSLLEDMDGLV 1018
DB 961 WMIDSECRPPFRELVSFSSMARDPQFVVIQNEIDGAPASPLDSTFYRSLLEDMDGLV 1020
QY 1019 DAERYLVPOQGFPPDPAAGAMVHHRHSSSTRSGGDLTGLFSESEEARSPAPS 1078
DB 1021 DAERYLVPOQGFPPDPAAGAMVHHRHSSSTRSGGDLTGLFSESEEARSPAPS 1080
QY 1079 EGAGSDVFDGDLGKAAGLQSLPTHDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQE 1138
DB 1081 EGAGSDVFDGDLGKAAGLQSLPTHDPSFLQRYSEDPTVPLPSETDGYVAPLTCSPQE 1140
QY 1139 YVNOPTVPPPPSEPPPLPAAPACATLERATLSFGKGVYKDVAFAGAVNPEYLT 1198
DB 1141 YVNOSEVQPPPPPLPPEGPPLPVPAGATLERATLSFGKGVYKDVAFAGAVNPEYLT 1200
QY 1199 POGGAAPPPPPPAFSPAPNLYYVODDPERGAPSTFPGTPTAENPEYLGADVP 1255
DB 1201 PREGTAPPPPPPAFSPAPNLYYVODDSEEGPPPSNFGCTPTAENPEYLGADVP 1257

RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
ID ERB2_MESAU
AC 060553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_Taxid=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. G930 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.

 DR EMBL; D16295; BAA03801.1; -
 DR HSBP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_kinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; kinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_kinase; 1.
 DR SMART; SM00261; F0; 3.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR Transmembrane; PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 1254
 FT DOMAIN 22 652
 FT TRASNEMEM 653 675
 FT DOMAIN 676 1254
 FT DOMAIN 158 368
 FT DOMAIN 472 644
 FT DOMAIN 720 987
 FT NP_BIND 726 734
 FT BINDING 753 753
 FT ACT_SITE 845 845
 FT DISULFID 189 204
 FT DISULFID 189 212
 FT DISULFID 236 244
 FT DISULFID 240 252
 FT DISULFID 255 264
 FT DISULFID 268 295
 FT DISULFID 299 311
 FT DISULFID 315 331
 FT DISULFID 334 338
 FT DISULFID 511 520
 FT DISULFID 515 528
 FT DISULFID 541 540
 FT DISULFID 544 560
 FT DISULFID 563 576
 FT DISULFID 567 584
 FT DISULFID 587 596
 FT DISULFID 600 623
 FT DISULFID 626 634
 FT DISULFID 642 642
 FT MOD_RES 1139 1139
 FT MOD_RES 1247 1247
 FT CARBOHYD 68 68
 FT CARBOHYD 125 125
 FT CARBOHYD 187 187
 FT CARBOHYD 259 259
 FT CARBOHYD 530 530
 FT CARBOHYD 571 571
 FT CARBOHYD 629 629
 FT VARIANT 658 658
 FT VARIANT 659 659
 FT SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;
 Query March 85.5%; Score 5841.5; DB 1; Length 1254;
 Best Local Similarity 85.7%; Pred. No. 2.6e-301;
 Matches 1076; Conservative 57; Mismatches 121; Indels 1; Gaps 1;

QY 61 ELTYLPNTASLSFQDIOEVGYVLIANNOYVQPLQRLRVRGTOQFEEDYALAVLDNG 120
 DB 61 ELTYLPNTASLSFQDIOEVGYVLIANNOYVQPLQRLRVRGTOQFEEDYALAVLDNR 120
 QY 121 DPLNNTPTVPGASGGELRELRLSTELKRGVLIQSNPOLCYODTLLMKDIFKKNQOLA 180
 DB 121 DPLNNTPTVPGASGGELRELRLSTELKRGVLIQSNPOLCYODTLLMKDIFKKNQOLA 180
 QY 121 DPLNNTPTVPGASGGELRELRLSTELKRGVLIQSNPOLCYODTLLMKDIFKKNQOLA 180
 DB 121 DPLNNTPTVPGASGGELRELRLSTELKRGVLIQSNPOLCYODTLLMKDIFKKNQOLA 180
 QY 181 LTLIDTRSRACHPCSPMCKSPGSSSESCOSLTETVACAGGAKRGLPTCCHEQC 240
 DB 181 LTLIDTRSRACHPCSPMCKSPGSSSESCOSLTETVACAGGAKRGLPTCCHEQC 240
 QY 181 LTLIDTRSRACHPCSPMCKSPGSSSESCOSLTETVACAGGAKRGLPTCCHEQC 240
 DB 181 LTLIDTRSRACHPCSPMCKSPGSSSESCOSLTETVACAGGAKRGLPTCCHEQC 240
 QY 241 AAGCTGPRGSDCLALHFNHSGICELHCPALVTYNTDFFSMPNREGRYTFCASCVTACP 300
 DB 241 AAGCTGPRGSDCLALHFNHSGICELHCPALVTYNTDFFSMPNREGRYTFCASCVTACP 300
 QY 241 AAGCTGPRGSDCLALHFNHSGICELHCPALVTYNTDFFSMPNREGRYTFCASCVTACP 300
 DB 241 AAGCTGPRGSDCLALHFNHSGICELHCPALVTYNTDFFSMPNREGRYTFCASCVTACP 300
 QY 301 YNVLSTVGSTLVCPLHNOGYVAEDTQCEKSKRCARVCYGLNQYIKANSKFTGIT 360
 DB 301 YNVLSTVGSTLVCPLHNOGYVAEDTQCEKSKRCARVCYGLNQYIKANSKFTGIT 360
 QY 301 YNVLSTVGSTLVCPLHNOGYVAEDTQCEKSKRCARVCYGLNQYIKANSKFTGIT 360
 DB 301 YNVLSTVGSTLVCPLHNOGYVAEDTQCEKSKRCARVCYGLNQYIKANSKFTGIT 360
 QY 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVQVEFTEITGYLYISAMPDQLP 420
 DB 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVQVEFTEITGYLYISAMPDQLP 420
 QY 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVQVEFTEITGYLYISAMPDQLP 420
 DB 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVQVEFTEITGYLYISAMPDQLP 420
 QY 421 DLSYFQNLQVIRGRILHNGAYSLTGLGSLWGLSRLBELSGGLALIHNTHLCFVHTV 480
 DB 421 DLSYFQNLQVIRGRILHNGAYSLTGLGSLWGLSRLBELSGGLALIHNTHLCFVHTV 480
 QY 421 DLSYFQNLQVIRGRILHNGAYSLTGLGSLWGLSRLBELSGGLALIHNTHLCFVHTV 480
 DB 421 DLSYFQNLQVIRGRILHNGAYSLTGLGSLWGLSRLBELSGGLALIHNTHLCFVHTV 480
 QY 481 PMDQLFENPHQALHTNRPEDCEVGGGLACHOLCARGHGWGPGPTQCVNCSQFLRQEC 540
 DB 481 PMDQLFENPHQALHTNRPEDCEVGGGLACHOLCARGHGWGPGPTQCVNCSQFLRQEC 540
 QY 481 PMDQLFENPHQALHTNRPEDCEVGGGLACHOLCARGHGWGPGPTQCVNCSQFLRQEC 540
 DB 481 PMDQLFENPHQALHTNRPEDCEVGGGLACHOLCARGHGWGPGPTQCVNCSQFLRQEC 540
 QY 541 VEECKVYQGLPREYVNAHCLPCHPECOPNGSVTCGPEADQCVACAHYKDPFCVYAR 600
 DB 541 VEECKVYQGLPREYVNAHCLPCHPECOPNGSVTCGPEADQCVACAHYKDPFCVYAR 600
 QY 541 VEECKVYQGLPREYVNAHCLPCHPECOPNGSVTCGPEADQCVACAHYKDPFCVYAR 600
 DB 541 VEECKVYQGLPREYVNAHCLPCHPECOPNGSVTCGPEADQCVACAHYKDPFCVYAR 600
 QY 601 PSQVYKPLSTMPIMKPPDEGACQPCINCTHSCVDLDKGCFAEGQASPLTJIFNNFTV 660
 DB 601 PSQVYKPLSTMPIMKPPDEGACQPCINCTHSCVDLDKGCFAEGQASPLTJIFNNFTV 660
 QY 601 PSQVYKPLSTMPIMKPPDEGACQPCINCTHSCVDLDKGCFAEGQASPLTJIFNNFTV 660
 DB 601 PSQVYKPLSTMPIMKPPDEGACQPCINCTHSCVDLDKGCFAEGQASPLTJIFNNFTV 660
 QY 661 SFWLRYPKVASHLEKRRQOKIRKTYMRLTJETELVEPLTPSGAMPNOQMOMLILETEL 720
 DB 661 SFWLRYPKVASHLEKRRQOKIRKTYMRLTJETELVEPLTPSGAMPNOQMOMLILETEL 720
 QY 661 SFWLRYPKVASHLEKRRQOKIRKTYMRLTJETELVEPLTPSGAMPNOQMOMLILETEL 720
 DB 661 SFWLRYPKVASHLEKRRQOKIRKTYMRLTJETELVEPLTPSGAMPNOQMOMLILETEL 720
 QY 721 RRYKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLEBNTSPKANKIILDEAYVAGVSP 780
 DB 721 RRYKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLEBNTSPKANKIILDEAYVAGVSP 780
 QY 721 RRYKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLEBNTSPKANKIILDEAYVAGVSP 780
 DB 721 RRYKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLEBNTSPKANKIILDEAYVAGVSP 780
 QY 781 YVSRLLGICLSTVQVLTQMPYGCILLDHYREHGRGLSGODLNMVCQIAKQMSYLEDVR 840
 DB 781 YVSRLLGICLSTVQVLTQMPYGCILLDHYREHGRGLSGODLNMVCQIAKQMSYLEDVR 840
 QY 781 YVSRLLGICLSTVQVLTQMPYGCILLDHYREHGRGLSGODLNMVCQIAKQMSYLEDVR 840
 DB 781 YVSRLLGICLSTVQVLTQMPYGCILLDHYREHGRGLSGODLNMVCQIAKQMSYLEDVR 840
 QY 841 LVHRDLAARNVLVKSNHVKITDFGLARLLIDETEHADGGKVPKIMMALESILRRFT 900
 DB 841 LVHRDLAARNVLVKSNHVKITDFGLARLLIDETEHADGGKVPKIMMALESILRRFT 900
 QY 841 LVHRDLAARNVLVKSNHVKITDFGLARLLIDETEHADGGKVPKIMMALESILRRFT 900
 DB 841 LVHRDLAARNVLVKSNHVKITDFGLARLLIDETEHADGGKVPKIMMALESILRRFT 900
 QY 901 HQSDVWSYGVYTWELTFGAKPYDGIIPAREIPDLLEKGRLLPQPPCTIDVYIMVYKCM 960
 DB 901 HQSDVWSYGVYTWELTFGAKPYDGIIPAREIPDLLEKGRLLPQPPCTIDVYIMVYKCM 960
 QY 901 HQSDVWSYGVYTWELTFGAKPYDGIIPAREIPDLLEKGRLLPQPPCTIDVYIMVYKCM 960
 DB 901 HQSDVWSYGVYTWELTFGAKPYDGIIPAREIPDLLEKGRLLPQPPCTIDVYIMVYKCM 960
 QY 961 IDESCRPRPRELVSEFSRMAPDPORFVUQNEDLGASPLDSTFTRSLLEDMDGLVNA 1020
 DB 961 IDESCRPRPRELVSEFSRMAPDPORFVUQNEDLGASPLDSTFTRSLLEDMDGLVNA 1020
 QY 961 IDESCRPRPRELVSEFSRMAPDPORFVUQNEDLGASPLDSTFTRSLLEDMDGLVNA 1020
 DB 961 IDESCRPRPRELVSEFSRMAPDPORFVUQNEDLGASPLDSTFTRSLLEDMDGLVNA 1020
 QY 1021 EBYLYVQGGFCDDPAPAGGVMVHHRHSSSTRSGGDLTLGLBPESEEPAPSLAPSEGG 1080
 DB 1021 EBYLYVQGGFCDDPAPAGGVMVHHRHSSSTRSGGDLTLGLBPESEEPAPSLAPSEGG 1080
 QY 1021 EBYLYVQGGFCDDPAPAGGVMVHHRHSSSTRSGGDLTLGLBPESEEPAPSLAPSEGG 1080
 DB 1021 EBYLYVQGGFCDDPAPAGGVMVHHRHSSSTRSGGDLTLGLBPESEEPAPSLAPSEGG 1080
 QY 1081 AGSDVFDGDLGGAAGKQSLPTHPBSPLOKQSEDPYVPLBETGTGYAFLTCSQPEYV 1140
 DB 1081 AGSDVFDGDLGGAAGKQSLPTHPBSPLOKQSEDPYVPLBETGTGYAFLTCSQPEYV 1140

QY 1141 NOPDVPPPPSPPEGLPAPAPGATLEERAKTSLSPGKGVVQVAFGAVENPEYLTPO 1200
 DB 1141 NOPDVPPPPSPPEGLPAPAPGATLEERAKTSLSPGKGVVQVAFGAVENPEYLTPO 1200
 QY 1201 GGAAQPPPPPPSPPAFDPNLYYWDQDPPEPGAPSTFKGPTAENPEYLGDPVY 1255
 DB 1201 GGAAQPPPPPPSPPAFDPNLYYWDQDPPEPGAPSTFKGPTAENPEYLGDPVY 1255
 RESULT 4
 EGFR_HUMAN
 ID EGFR_HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9JMG5; Q92795; Q00732;
 AC Q00688; Q98252; Q9G2C9; Q9GZX1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Varden Y., Liberman T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells."
 RT Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7664368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta."
 RT Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter U.L., Maible N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor."
 RT Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer."
 RT Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 3 AND 4).
 RX TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;
 RA Reiter U.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms."
 RT Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter U.L., Thredgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Maible N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor."
 RT Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=3226261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verna I.M., Gill G.N., Rosenfield M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells."
 RT Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells."
 RT Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells."
 RT Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription."
 RT Oncogene Res. 1:375-386(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis."
 RT J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene."
 RT Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA."
 RT Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=99278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honneger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor."
 RT J. Biol. Chem. 264:10667-10671(1989).
 RN [16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=66398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RL ovary fibroblasts.";
 RL Growth Factors 13:121-132(1996).
 RN (17)
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim U.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor.";
 RL J. Biochem. 127:65-72(2000).
 RN (18)
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Otake M., Inagaki P., Iak I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor.";
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN (19)
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens.";
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor. Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;
 DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;
 Query Match 45.5%; Score 3111; DB 1; Length 1210;
 Best Local Similarity 49.5%; Pred No. 4,3e-157;
 Matches 625; Conservative 172; Mismatches 364; Indels 102; Gaps 21;

QY ASLSFLDIOIEVOGVVLIANQVROVPLORRIYRGTLPEEDYALAVLJONGDPLNNTTP 128
 DB 74 YDLSFLKIQIAGVAVVLIANTVERIPENLIQIRGMYIENSTALAVLSYD----- 126
 QY 129 VTGASPGGLRELOLRSLTEILKGVLLIQRPOLCYODTILMKDI FHNQNALATLIDTNR 188
 DB 127 ---ANKTGKELPRNLOELHGAVERSPNPALCNVESIQWRDIVSSDPLSMMDPQNH 183
 QY 189 SRACHPSPMCKSKRCWCESEDECOILTRYCAGGA-RCKGFLPTDCCEQCAAGCTGP 247
 DB 184 LGSQCKDPSCPNCSGWAGEENQKLTIKTICACQCSGRKRGKSPSDCHQCAAGCTGP 243
 QY 248 KHSQDLALHNHNSGICELHCPALVYNTDTFESMPREGYTFGASCVTACPYNYLTD 307
 DB 244 RESDCLVCRKRDATCKDCPFLMLNPTTYQMDVNPBGKYSGATCVKCKPNNVYTD 303
 QY 308 VGSCTLVCPHNOVVTMEDGTORCEKSKRCARCYGLGQYIKANSKFTGITELE-PAG 366
 DB 304 HGSVCRAAGADSYEM-BEDGVKCKKCEGCRKCNIGIGIEFK-DLSINATNIKHFKN 361
 QY 367 CKKIFGSLAFPSFSDGDPASNTAPLQPELOVETLEETGYLYISAMPDLSLVSFQ 426
 DB 362 CTISISDLHLIPVAFRGDSTHTPPLDPQEDLIKTVKEITGFLIQAMPENRDTHAFE 421
 QY 427 NLQVIRGRILHNGAVSULTLOGISWLGSLSLRELGSGLALIHNTLHCFVHTVPQDLF 486
 DB 422 NLEIRRTKQHQGFSLAVVSLNITSGLNSLVEISGDVYISGNKMLCVANTINWKLFL 481
 QY 487 RNPQALHTANRPEDCEVBEGLACHQLCARHGCKWGPPTQCVNCSQFLRGQCEVECRV 546
 DB 482 GTSGQKTKIISNRKNSCKATQVCNALCPSEGWCPEPRDVCERNVSGRECVDCNLT 541
 QY 547 LQGLPREYVNAARCLPCHPFCPONGSVTQPGBADQVCAHYKDPFVACPSGVK 606
 DB 542 LEBGRRFRVENSICQHPBCLPQAMITLTGSGPNDICCAHYIDBPCVKTCPAGVMG 601
 QY 607 DLSMPIMKFPDEEGACQPCPINCTSCVLDLDDKGPAGRASPLTSIFNNFTVSFLRV 666
 DB 602 ENNTLT-VWKYADAGHVCHLHPNCTYGTCTPGLEGCTNGPKLP-SIANGVALLLL 658
 QY 667 PKVSASHLEBRROKRIKRYMRRLLOETELVEPLTSGAMPNQAORILKETRYKKVL 726
 DB 659 VVALGIGLFMRHRHIVKRLRLQERELVEPLTSGAMPNALLRIKETEFKKIKVL 718
 QY 727 GSGAFGVYKGIWIPDGENYKIPVAIKVIRENTSPKANKIILDEAYVMAGVSPYRRL 786
 DB 719 GSGAFGVYKGIWIPGEKXKIPVAIKELREATSPKANKIILDEAYVMASVDNPHVCRLL 778
 QY 787 GILTSIVOLVQLMRYGCLLDHVENRGRLSGQDLINWCMQIAKMSYLEDVRLVHRL 846
 DB 779 GILTSIVOLITLMPFGCLLDYREHKONIGQYLLNWCQIAKMSYLEDRLLVHRL 838
 QY 847 AARNVYKSPNHKINDPGLARLDIDETEVHADGKVPKYMALSSILRRTHOSDVW 906
 DB 839 AARNVYKTPQHKITDPGLAKLGAEBKYNHAGGKVPKYMALSSILHRIYTHOSDVW 898
 QY 907 SYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGRLEPOPICTIDVYIMVKKCMIMSECR 966
 DB 899 SYGVTVWELMTFGSKPYDGIIPAREISILSEKGRLEPOPICTIDVYIMVKKCMIMDADR 958
 QY 967 PRFRELVEFSRMAPRQRFVYIQ-NEDIGPASPLDSTRYSLLEDDMDGLVDAAEYLV 1025
 DB 959 PRFRELITERSKAPRQRYLYIQGERHNLSPDTSNEYRALMDEEDMDVDVDAEYDI 1018
 QY 1026 PQGPFCDPAPAGAGVHHRSSSTRSGGDLTLGPSEBEARSPLASSEGAGSDV 1085
 DB 1019 PQGPF-----SSPSTRITLSSLSKTSN- 1043
 QY 1086 PQGDLGMAKAGIOSIPTHDPBPLQRYSDPVPVLPSET--DGYVAPLTCSPQEPYVNP 1143
 DB 1044 -NSTVACIDRNLGOSCPIKEKESFLQRYSSDPTGALTEDIDDTFL-----VYPEYING- 1095

QY 1144 DVPFPSPSEGEPLPAPAPAGATLTERAKTLSPGKNGVXVDFAGAVENPEYL-TPQGG 1202
 DB 1096 SVKCPAGSVQNPVYHNPCLNP-----APSRDPHYGD--PSTAVGNPEYLTVQ-- 1143
 QY 1203 AAPQHPHPAPFAPFNLYYMQ-----DP-----PERGAPSTFGTGTAEHP 1246
 DB 1144 -----PTCVNSTFDSPAHMAOKGSHQISLDNPDYQDFFPEAKKPNKIFKGS-TAENA 1195
 QY 1247 EYL 1249
 DB 1196 EYL 1198
 RESULT 5
 EGFR_MOUSE
 ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
 AC 001279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RX MEDLINE=93026370; PubMed=1408137;
 RA Avivi A., Skorecki K., Yaron A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 (heK/KGFR) gene."
 RL Oncogene 7:1197-11962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; and CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; PubMed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Day S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 in mouse blastocysts during delayed implantation."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Hibbs M.L.;
 RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; PubMed=8125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 receptor tyrosine kinase."
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=91232866; PubMed=2030916;
 RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 binding site."
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Bisinger D.P., Serrero G.;
 RT Submitted (Jun-1992) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X78987; CA55587.1; -
 CC EMBL; U03425; AA17899.1; -
 CC EMBL; X59698; CAA4219.1; -
 CC EMBL; L06864; AAA53029.1; -
 CC EMBL; Z12608; CAA78249.1; -
 CC HSSP; P11362; IFGK.
 CC MGD; MGI:95294; EGF.
 CC InterPro; IPR000494; EGFR_L domain.
 CC InterPro; IPR000719; Euk_Pkinase.
 CC InterPro; IPR002174; Puriin-like.
 CC InterPro; IPR001245; Tyr_Pkinase.
 CC Pfam; PF00069; Pkinase; 1.
 CC Pfam; PF00757; Puriin-like; 1.
 CC Pfam; PFO1030; Recep_L domain; 2.
 CC ProDom; PD000001; Euk_Pkinase; 1.
 CC SMART; SM00261; FU; 3.
 CC SMART; SM00219; TYKc; 1.
 CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 CC PROSITE; PSS0011; PROTEIN KINASE DOM; 1.
 CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 CC FT SIGNAL 1 24
 CC FT CHAIN 25 1210
 CC FT DOMAIN 25 647
 CC FT TRANSMEM 648 670
 CC FT DOMAIN 671 1210
 CC FT REPEAT 75 300
 CC FT REPEAT 390 600
 CC FT DOMAIN 1028 1071
 CC FT NP_BIND 714 981
 CC FT BINDING 720 728
 CC FT ACT_SITE 747 747
 CC FT DISTULFID 839 839
 CC FT DISTULFID 190 199
 CC FT DISTULFID 194 207
 CC FT DISTULFID 215 223
 CC FT DISTULFID 219 231
 CC FT DISTULFID 232 240
 CC FT DISTULFID 236 248
 CC FT DISTULFID 251 260
 CC FT DISTULFID 264 291
 CC FT DISTULFID 295 307
 CC FT DISTULFID 311 326
 CC FT DISTULFID 329 333
 CC FT DISTULFID 332 340
 CC FT DISTULFID 510 523
 CC FT DISTULFID 526 535
 CC FT DISTULFID 539 555
 CC FT DISTULFID 558 571
 CC FT DISTULFID 582 591
 CC FT DISTULFID 595 617
 CC FT DISTULFID 620 628
 CC FT DISTULFID 624 636
 CC MOD_RES 680 680
 CC MOD_RES 1092 1092
 CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

FT	MOD_RES	1110	1110	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	MOD_RES	1112	1172	PHOSPHORYLATION (AUTO-)	(BY SIMILARITY)
FT	MOD_RES	1197	1197	PHOSPHORYLATION (AUTO-	MAJOR SITE)
FT				(BY SIMILARITY)	
FT	CARBOHYD	128	128	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	175	175	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	196	196	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	352	352	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	413	413	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	444	444	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	528	528	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	568	568	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	603	603	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CARBOHYD	623	623	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT	CONFLICT	19	19	C -> S (IN REF. 2)	
FT	CONFLICT	539	539	C -> W (IN REF. 5)	
FT	CONFLICT	991	991	L -> F (IN REF. 4)	
FT	CONFLICT	1116	1117	HP -> DR (IN REF. 6)	
SO	SEQUENCE	1210 AA;	134853 MM;	690E820D46DF22D2P5	CRC64;

Query Match	Score 3081;	DB 1:	Length 1210;
Best Local Similarity	49.3%;	Pred. No. 1.6e-155;	
Matches 627;	Conservative 164;	Mismatches 371;	Indels 110; Gaps 23

QY	11	LLALLPPGAA--SYOVCTGIDMKRLPLASPEITHIDMKRLHVXGQVQVQGNLEIYLPFN	68
Db	14	LITLALCAAGALAEKKVCOGISTNRLTQLOTFEDHFLSLQRMNNECVLGNLEIYYQNRN	73
QY	69	ASLSLQDIQOEYQYVLIAHNOVQPLQRLRIIVEGTOLFEDNVALAVLDNGDPINNTPP	128
Db	74	YDLSLTKTIOEYAGVYLLALNIVERIPLENLOIIGNALYENTVALIISN-----	124
QY	129	VTGASPGGLREIQLSLTEILKGGVLIQNPQLQODIILMKDI---FHKNQLATLLI	184
Db	125	-YGNRTQLREIPLMNLOEILIGAVFSNNPLLCNMDDTIQMDIYQNVMSMSMDL---	180
QY	185	DTNRSRACHPCSPMCKSGRCWGESSEDCQSLRTVCAAGCA-RCMPLPTDCCHQCAAG	243
Db	181	-QSPHSRCPKCDPSCPCNCSGCGGEENCQKLTIKICAOQCSHRGCRGSPDCHQCAAG	239
QY	244	CTGHNSHCLACLHNHNSGICELHFPALVTYNTDFEEMPNNEGTYTGASGVYACPNY	303
Db	240	CTGRRESCLVQCFQDEATCKDTCPMLNPTTYQMDVNBEGKISGATCVKCKPNY	299
QY	304	LSTDVGSCTVCPJHNOETVADGTQRCCKSKPCARVYGLQMYIYANSKFIGITBLE	363
Db	300	VVTHGSCVRAACGPDYEVV-EEDGIRKCKCDGPRKVCNGHIGIEFX-DTISINATIK	357
QY	364	-FAQCKKIFGSLAFELPESFGDDGPAINTPLQEOQVETLEIETIGYIYISAMPDLPDL	422
Db	358	HFKICTAISGHLILPVAFKDSFRTPLDLPRELEILIKYKELTGFLLIQAMPDNDTL	417
QY	423	SVFQNLQVIRGRILHNGAYSLTLQGLISWLGRLSLRLSGSLALIHNTHTLCHFHYPM	482
Db	418	HAFENLEIIRRTKQHGQFSLAVVGLNITSLGLSLKEISDGYIISNRMLCVANTINM	477
QY	483	DQLRNHPQALHTNANPEDECBVEGLACHQLCARGHWGGGPPQCVNCSQPLRGQCEVE	542
Db	478	KQLGTBPQTKIIMNRPKEKCKKAVHVCNPLCSSEGCGWGPBPDCVSCQNVSGRECEVE	537
QY	543	ECRYLQGLPREYVARNHQLPCHPCEOPONGSVTFGFPADQCVACAHYKDPFCVACRPS	602
Db	538	KCNILIEGEPREFVENSECIOCHPECLPQAMNITCTGREGPNCIOCAHIIDBPHCVKCPA	597
QY	603	GVRKDLVMPRIWKRPDEGACQPCPINTHSCVLDLDKGCRAEQRASPLNISFNNFTVSF	662
Db	598	GIMENNTL-VMKYADANNVCHLCHACTYGCAGPGLQGCWVMSGPKRIPSIATIGIYGL	656
QY	663	WLRPVKSASHLEKRRQOKIRKYTMRLLOETLEVEPLTSGAMPNOQMILKETELRK	722
Db	657	LFIYVVALIGLIFMRRHIVAKTKTLRRLOREILVEPLTPSGEAPNQAHRLIKETEFK	716
QY	723	VKVLGSGAFGVYVIGIMIPDGENVKYPAIKVURENTSPRANKILDEAYVACVSGPDV	782

DL	717	IKVLGSGAGFTVYKGLMIPEGEKVKIPVAIKELBAPATPKANKELIDEAYVASVDNRHV	776
QY	783	SRLLGICLTSTVOLVTOQLMPYGGCLLDHRENRGRGLGSDLLNMCQIHKGSYLEDEVLY	842
Db	777	CRLLGICLTSTVOLITOLMPYGCCLLDVYREKMDNGSQYLLNMCQIHKGNVYLEDRLY	836
QY	843	HRDLAARVNLVKSPEHNVKVTITDFGLARLLDIDETETVHADGKVPKIMALESTLRPRFTHQ	902
Db	837	HRDLAARVNLVKTQHQVKTITDFGLAKLGAKEKYHAGKVPKIMALESTLRHTYHQ	896
QY	903	SDVMSYGVTVWELMTFGAKPYDGI.PAREIPDLLEKGERL.PQPPICITIDVYIMVXCMMD	962
Db	897	SDVMSYGVTVWELMTFGSKPYDGI.PASDISILEKGERLPQPPICITIDVYIMVXCMMD	956
QY	963	SECRPREELVSEFSRMARDPQRVVVQ-NEDLGPASLDSFTFRSLLEDODMGDLVAAE	1022
Db	957	ADSRKPEFELLLESKKAARDPQRYLVQGDGRMLPSTDSNFYRALMDEEDMEDVAD	1016
QY	1022	EYLVPOQGFPCPDPAFGAGVWHRHSSSTRSGCDLTGLLEPSEEAPEPSPLAPSEGA	1083
Db	1017	EYLVPOQGF-----NSPST-----SPTPLSLISA	1042
QY	1082	GSDVFDGDLGMAKAGLQSLPTDHPSPLOYSEDPYPLPSET--DGVAPLTSCPOEY	1139
Db	1043	TSN-----NSTVACINRNGSCVKKEDAFLORYSSDPTGAVMEDNIDDAFL-----PVPEY	1092
QY	1140	VNCPDVPKPPSPREGPLPARPAGATLERAKTLSPKNGVYKVFAGVAPSEYL-T	1198
Db	1093	VNO-SVPRKPPGSGVNYHNPQLP-----APGDLHYQN--PHSNVNGPEYNT	1141
QY	1199	PQGGAPQHPHPAPSPAFDNLVYMDQ-----DP-----PERGAPSTFKGPT	1242
Db	1142	AQ-----FTLSSGGSNPALMIQKSHQSLMDPDVQDFFPKETKNGIFKG-PT	1191
QY	1243	AENPEYGLDVP	1254
Db	1192	AENAEYLRVAP	1203

RESULT 6

ID	ERB4_HUMAN	STANDARD;	PRT;	1308 AA.
AC	Q15303;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)			
DE	(p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).			
GN	ERBB4 OR HER4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
NC	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM JM-A).			
RP	TISSUE=Fetal brain;			
RC	MEDLINE=97476287; PubMed=8383326;			
RX	MEDLINE=93189574; PubMed=8383326;			
RA	Ploewman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,			
RA	Foy L., Neubauer M.G., Shoyab M.,			
RT	"ligand-specific activation of HER4/p180erbB4, a fourth member of the			
RT	epidermal growth factor receptor family."			
RT	Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).			
RL	[2]			
RN	SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).			
RP	TISSUE=Fetal brain;			
RC	MEDLINE=97476287; PubMed=9334263;			
RX	MEDLINE=97476287; PubMed=9334263;			
RA	Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Ploewman G.D.,			
RA	Klaegsburn M.,			
RT	"a novel juxtamembrane domain isoform of HER4/ErBb4. Isoform-specific			
RT	tissue distribution and differential processing in response to			
RT	phorbol ester."			
RL	J. Biol. Chem. 272:26761-26768(1997).			

CC - FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NR-
 CC 2, NR-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- α , AND AMPHIREGULIN.
 CC
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC
 CC - SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS, JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC
 CC - TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: L07868; AAB59446.1; -
 CC HSSP: P1362; 1FGK.
 CC Gene: HGNC:3432; ERBB4.
 CC MIM: 600543; -
 CC
 CC InterPro: IPR000494; EGFR_L_domain.
 CC InterPro: IPR000719; Euk_kinase.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR001245; Tyr_kinase.
 CC InterPro: IPR004019; YLP_motif.
 CC Pfam: PF00069; kinase; 1.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF01030; Recep_L_domain; 2.
 CC Pfam: PF02757; YLP_2_kinase; 1.
 CC ProDom: PD000001; Euk_kinase; 1.
 CC SMART: SM00261; FU; 4.
 CC SMART: SM00219; TYKc; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane, Glycoprotein, Multigene family; Receptor; Signal;
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC Alternative splicing.
 CC
 CC FT SIGNAL 1 25
 CC FT CHAIN 26 1308
 CC FT DOMAIN 26 651
 CC FT TRANSMEM 652 675
 CC FT DOMAIN 676 1308
 CC FT DOMAIN 186 334
 CC FT DOMAIN 496 633
 CC FT DOMAIN 718 985
 CC FT NP_BIND 724 732
 CC FT BINDING 751 751
 CC FT ACT_SITE 843 843
 CC FT DISULFID 189 197
 CC FT DISULFID 193 205
 CC FT DISULFID 213 221
 CC FT DISULFID 217 229
 CC FT DISULFID 230 238
 CC FT DISULFID 234 246
 CC FT DISULFID 249 258
 CC FT DISULFID 262 289
 CC FT DISULFID 293 304

FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 626 648 NPSHSCIYPTMGHSTLPOHA -> IGSSIEDICGLMD (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 58AEB0965D86761 CRC64;
 Query Match 42.7%; Score 2920.5; DB 1; Length 1308;
 Best Local Similarity 44.6%; Pred. No. 5,4e-147;
 Matches 604; Conservative 178; Mismatch 384; Indels 187; Gaps 26;
 QY 9 MGLLLALLPPGAA-----STQVCTGTOMKALPASPETHLMDLHLYGCGVQGNLELY 64
 DB 8 WWSVSLVLAAGTVPDSQSVACGTENKLSLSLDEQYALRKYENCCEVWGNLEITS 67
 65 LPTNASISFLDIDIEGVYLLAHNVQVPLRLRIVRSTOLFEDNVYALAVLNDGDLN 124
 DB 68 IEHNRDLSFLRSVREYVYVLAHQRYVPLENLRIRIIRKLYEDRYALALFLNVRKDQ 127
 QY 125 NTTPVTGASPEGLEBELDRLSITELKGVLIQRNPOLCYODTILMKIIFKNNOLATLI 184
 DB 128 NF-----GLQELGLKULTELLNGVYVDQKFLCYADTIHQDIVNPNPNSLTLV 178
 QY 185 DYNRSRACHPSPWCKGSRGCGESSQCSUTRTVCAGG-ARCKGPLPTDCHEQCAAG 243
 DB 179 STNSSGCGRCHSKCTG-RCMGPTENHCQILTRIVCAQCDGRCTGYVSDCHRECAAG 237
 QY 244 CTGPKHSDCLALHFNHSGICEHLCPALVTYNTDTFESMPNPEGRYTFGASCYTA CPYNY 303
 DB 238 CSGPKDNDPCFACMNFNSGACVTCQPTFYVNPFTPLEHNFNAKYTGAFVCVKKCPHNF 297
 QY 304 LSTVSGCTLVCPALHNEVAEDSTQCEKSCRCARVCGIQMAYKANSKFIQLELE 363
 DB 298 V-VDSSCVRACPSSKREV-BENGIKMKCKCTDIPCACGIGISMSQTVDSNIDK 355
 QY 364 FAGCKKIFGSLAPFESPDPAASNTAPLPBOLQVETLEITGYLYISAMPDSDLDS 423
 DB 356 FINCTKINGNLIPLVTGIGHDPYVAIDPEKLNVRVRYREITGFLINSWPNNMDFS 415
 QY 424 VFQULQVIRGILNNGAVSLTQLGISMGLSRLSGSLALIHNTLHCFHYTPMD 483
 DB 416 VFSNLVTIGGVVLSGLLILKQGLTSIQFSLKISAGNIYITNSNLCTHYHTLNT 475
 QY 484 QLFPNPHQALLHTANRPEDECVGEGLAQHOLCARGHCGPGPTCVVACFLRGOECVEE 543
 DB 476 TLFTINQRIVIRNRGAENCTABGWCNHLCSDDGWMGPGPODCLSCRFSGRICTIES 535
 QY 544 CRVIGGLPREYVNAHCLPCHPBCOP-QNGSVTCFGEADQCCVACAHYKPPFVACPS 602

Db 536 CNYDGFREFRNGSTVECDPOCEKEDGLTCHGPDNCTKCSFKGPNCEKCPD 595
 Qy 603 GVXPDLSYPMWKEPDEGACQCPINCSTHSCVDLDDKGPAPGASPLTSIF-----NNF 658
 Db 596 GLOQANSF--IFKYADDRECHPCPNCTGCGNPTSHDC-----IYPMTHGS 642
 Qy 659 TVSFMLFVPRKVSAS-----HLEKRRQCKIKRYMRRLQTELVPELTPS 703
 Db 643 TLPOHARTPLIAAGVIGLFLYIVGLTFVAVYRKRKIKKKRALRRL-ETELVEPLTPS 701
 Qy 704 GAMPNQAMRILKETELRYKVGSGAFYVYKGIWIPDGENYKIPAIIVLENTSPKA 763
 Db 702 GTANQOLRLKETELRYKVGSGAFYVYKGIWIPDGENYKIPAIIVLENTSPKA 761
 Qy 764 NKEILDEAVYVAGVGSPPVSRBLGICLTSTVQVLTQMPYGLLDHVRNRRGLSGDLL 823
 Db 762 NVEEMDEALIMASNDHRLVRLGLVCSPTIQLTQLMPHGLLEVEHEKMDIGSOLL 821
 Qy 824 NMCWQIAKMSYLEDVRLVRLDAAARVLYKSPHVKITDPGLARLIDIDETRYHADGK 883
 Db 822 NMCWQIAKMSYLEDVRLVRLDAAARVLYKSPHVKITDPGLARLIDIDETRYHADGK 881
 Qy 884 VPIKMALESILRRFTHQSDVMSYGVTVMLTFGAKPYDGIIPAREIPDLLEKGERLPQ 943
 Db 882 MPKMALECIHYKFTHQSDVMSYGVTVMLTFGAKPYDGIIPAREIPDLLEKGERLPQ 941
 Qy 944 PPICTIVYMWKCMWIDSECRPRELYSEFSRMDQRFVYVIONED-LGAPSLDS 1002
 Db 942 PPICTIVYMWKCMWIDSECRPRELYSEFSRMDQRFVYVIONED-LGAPSLDS 1001
 Qy 1003 TFRSLDEDDMGDLVAEEYLPQCGFCPPDPAAGAVHHRRSSSTRSGGDLTLG 1062
 Db 1002 KFFGNLDEEDLEMMDAEYLP-QAFNIPPP-----YTSRARIIDSRS-----EIG 1049
 Qy 1063 LEPSEEEAPPS-----PLAP-SEGAGSDVFDGIGWG 1093
 Db 1050 HSPRPAYTPWGNQFVRRDGFAGAEQGVSVYRAPSTIEEAPVACATHEIPDDSCNG 1109
 Qy 1094 AAKLOSLPTHDSPLORYSEDPTVPLPS-----ETDGVVAPLTCSPQREYVNPQDVA 1146
 Db 1110 TLKKPVAPHVQSDSTORYSADPTVFAPERSPGLEDEEGYMPMDKPKQOYLNPVE-- 1167
 Qy 1147 PQPSPREGPLPAARPGATLERAKTLSPKNGVYKDVFAFGAVENPEYLLPQGGAAPO 1206
 Db 1168 -----ENPFVSR-----KNGDLQ-----ALDPEYHNASNG--- 1194
 Qy 1207 PHPPA-----FSAFNDLYWDDPPBERGA 1232
 Db 1195 --PPKADDEVYNEPLVNTANTLGAKEYLKNMILSNPERAKKAFDNPDMNSLPFRST 1252
 Qy 1233 --PPSTFKGTP-----AENPEYL 1249
 Db 1253 LQHDVLYQEYSTKYFYKQNGRIRIYVANEPEYL 1285

RA Marchionni M.A., Kelly R.A.;
 RT "Neuregulin promotes survival and growth of cardiac myocytes."
 RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
 RT ventricular myocytes."
 RL J. Biol. Chem. 273:10261-10269(1998).
 RN [2]
 RP SEQUENCE OF 848-901 FROM N.A.
 RC TISSUE=Sciatic nerve;
 RX MEDLINE=9122560; PubMed=2025425;
 RA Lai C., Lemke G.;
 RT "An expanded family of protein-tyrosine kinase genes differentially
 RT expressed in the vertebrate nervous system.";
 RL Neuron 6:691-704(1991).
 RN [3]
 RP SEQUENCE OF 1031-1198 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuregulin and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF- β , AND AMPHIREGULIN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
 CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
 CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
 CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
 CC HEART.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AF041838; AAC08899.1; -
 CC HSPR; U52531; AAC83051.1; -
 DR HSPR; P11362; 1FGK.
 DR InterPro: IPR000494; EGF_R_L domain.
 DR InterPro: IPR000719; Furin-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recp_L domain; 2.
 DR Pfam; PF02757; YLP_2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 4.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRANSMEM 652 675
 FT DOMAIN 676 1308
 FT CYTOPLASMIC (POTENTIAL).

```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA: 146957 MW: D944BB0996A08B41 CRC64;

```

Query Match 42.5%; Score 2906; DB 1; Length 1308;
 Best Local Similarity 44.5%; Pred. No. 3.1e-146;
 Matches 603; Conservative 185; Mismatches 392; Indels 174; Gaps 26;

```

QY 1 MEDA-ALCPWGLL--ALLPPGAATOVCTGTDMLRLPASPEHLMRLHYOGQOVQ 57
DB 1 MKLATGLWVGSLLVAVARTVQPSASOSVCGTENKLSLSLDLEQYRALRKYYENCEVYM 60
QY 58 GNELELYLPNALSFLQDIQEVQGYLLAHNVROVPLQELRIVRGTOLEFEDNALAVL 117
DB 61 GNELEISIEHNRDLSELRISIREVTGYLLVNLQFRYLPLENLRIIGTXLYEYRVALAF 120
QY 118 DNGDPLNNTTPTVGPAGSGRLRELQRLSLTEILKGYLIORNPOLCVQDITLWMDIHKXN 177
DB 121 LNVKXDNF-----GLQELGLKLTETILNGSVYVDONKFLCADITIHMODIVRANW 171
QY 178 QALATLIDNRSACHPCSPMCKSGRCWGESSEDCSLTRTVACAGC-ARCKGFLPTDCC 236
DB 172 PSMNTLVSTIGSGCGCHKSCG-RCWGPTEHNCQTLTRTVACAGCDDRCVGFVYSDCC 230
QY 237 HEGCAGSGCGPKSDCLACHFNHSGICELHCPALVTYNTDTPESNPNBGRYTFASCV 296
DB 231 HRECGAGSGGPKDTCACMKNENDSACVTCQCPQTVVNPFTTOLHNFNNAKTYGAFVCV 290
QY 297 TACPYNLTSDVSGCTLVCPILHNOEYTAEDGTORCEKCSKPCARVCYGLGMOYIKANSKF 356
DB 291 KKCPHNFV-VDSSSVACAPSSKMEV-EEINGIKCKPCTDIDCKKADGIGTGSIMSAQTV 348

```

```

QY 357 IGITELEPAGCKKIFGSLAFPLESPEDGDPASNTAPLOPEQLQVETLEEITGYLYISAMP 416
DB 349 DSNIDRKINCTKINGNIIPLVTGIDGPPVNAIDAPDEKLNVERTYREITGFLINQWMP 408
QY 417 DSLPLSTFQNLQVIRGILHNGAVSLTQSLGISWGLRSLRELSSLLAIHNTNLCF 476
DB 409 PNMDFSVFSLVYIGGRVLSGLSLILKQGGITSLQFOSLKEISAINIYITONSNCY 468
QY 477 VHTVMDLFPNPHQALHTANRPEDECEVSGLAGCHQACARHCGMCPPTQCVCNCSQPLR 536
DB 469 YHTIMTTLFSTVNDRIYIRDNRAENCTAAGMCMNLCSNDGCMGPRDCLSCRRSR 528
QY 537 GQCEVEECVLIQGLPREYVNAHCLPCHPECOP-QNSVTCFGEADQVCAHAYKDPF 595
DB 529 GKICIESCNLYDGEFEFENSGICVECDQCEKMEGDGLTCHGCPDCTCKSHFKDQPN 588
QY 596 CVARPPSVKRDLSYMPIMKPEPDEGACOPPCINCTSHCVLDDKCGCAEABASPLSIF 655
DB 589 CVERKCPDVLQANNSF--LFKYADODRECHPCHPNCCTQCGNGPTSHDC-----LY 635
QY 656 ---NFTVSFWLVPKVSAS-----HLEKROOKIRKRYTRRLLOETEL 696
DB 636 YPMTHGSHLPQHARTPLIAAGVIGGLFVLMALTPAVYARKSIKKRALRFL-ETEL 694
QY 657 VEPLTPSGAMPNQAOMRLKETEIRKRYVLSGAFYTKGIWIPDGENVKIPIVAKYLR 756
DB 695 VEPLTPSGTAPNQAOLRIKETELKRYVLSGAFYTKGIWIEGTVKIPVAKYLR 754
QY 757 ENTSPKAKEILDEAVYVAGSPYVSRLDICTSTVQVLTQMLPVCCLDHYHENG 816
DB 755 ETTGKXANVEEMDELINASVDHILVRLGVCSPITQVLTQMLPHOCLLEYHEND 814
QY 817 LGSODLNMVQIAKMSYLEVRLVHRDLAARVLYKSPNHVKITDFGLRLDIDETE 876
DB 815 IGSODLNMVQIAKMSYLEVRLVHRDLAARVLYKSPNHVKITDFGLRLDIDETE 874
QY 877 YHADGKXPIFMMALESLRBRFTHOSVMSYGTVMELMFGKPYGIPAREIDLE 936
DB 875 YNADGKXPIFMMALESLRBRFTHOSVMSYGTVMELMFGKPYGIPAREIDLE 934
QY 937 KGERLPQPICTIDVYIMVAKMIMDECEPRFRELSEFSRMAPDPORFVIONED-LG 995
DB 935 KGERLPQPICTIDVYIMVAKMIMDECEPRFRELSEFSRMAPDPORFVIONED-LG 994
QY 996 PASPLDSTFTYSLLEDMDGLVDAEELVYQOGFPCDP----- 1035
DB 995 LPSPDSTFTYSLLEDMDGLVDAEELVYQOGFPCDP----- 1033
QY 1036 ---APGAGMTHRRHRSSTRSGGDLTLGLEPSEEBAPRSLAPSEGAGSDVFDGLGM 1092
DB 1054 PAYTMSSGCVYVQDGFATOG---MMPYTAITSTIPAPVA--QATLMDMDSCN 1108
QY 1093 GAAGLQSLPHDPSPLQRYSEDPVLPJS-----ETDGYVAPLTCSPQPEYVNOVDV 1145
DB 1109 GLTRKFPVPHVQEOSSTQRYGADPLTFVAFERNPRAELDEBSYMTPMHCKPQEVLPVE- 1167
QY 1146 RPPQSPREGLPAPAPRAGATLBERAKTSLSPKNGVYKVFAPGAAVENPELTVQGGAP 1205
DB 1168 ---ENPFVSR-----KNGDLQ-----ALDPPEYHSSSG-- 1194
QY 1206 QPHPPA-----FSPAFLNLYWDDDPERG 1231
DB 1195 ---PEKADVEYNEPLVNTFTNALGNAEYKNSLSVBEKAKKAFQNDPDMNSLSPRS 1251
QY 1232 A--PPSTKGTPT-----AENEYL 1249
DB 1252 TLQHPDYLOEYSTKYFYKNGRIRPIYANENEYL 1285

```

RESULT 8
 XMRK_XIPMA
 ID_XMRK_XIPMA
 AC_P13388;
 STANDARD; PRT; 1167 AA.

Db 590 IPRCHGILGCDTL-IMKYADKMGCCOCHONOGSGPGLSGCRGDIYSHSLAVGL 648
 Qy 655 FNNFVSPFLRPXKXSASHLEKRRROOKIRKXTMRLLQETELVEFLTSGAMPNQAOKRI 714
 Db 649 VSGLLITVIVALLIV--LIRRRIRK-RKRTIRCLLDEKELVEFLTSGAMPNQAOKRI 704
 Qy 715 LKETELRKVKVLGSGAFIVYKGIIPDGENYKIPVAIKVLENTSPRANKEIDEAVVM 774
 Db 705 LKETEPRKORVLGSGAFIVYKGIIPDGENYKIPVAIKVLENTSPRANKEIDEAVVM 764
 Qy 775 AGVSPYVSRLLIGITCTITVQVLTQMLPFGCLLDVRENRKGLSGQDLNMCQIAKAMS 834
 Db 765 ASVDHPHVCRLIGITCTITVQVLTQMLPFGCLLDVRENRKGLSGQDLNMCQIAKAMS 824
 Qy 835 YLEEDVRLVHRDLAARVVLKSPNHYKITPFGALRLDIDETEHYADGGKVPKMALESI 894
 Db 825 YLEERVLVHRDLAARVVLKSPNHYKITPFGALRLDIDETEHYADGGKVPKMALESI 884
 Qy 895 LRRFTHQSDVMSYGVTVWELMTFQAKPYDGIIPAREIDPLEKGERLPQPICTIDVYMI 954
 Db 885 LQMTYHQSDVMSYGVTVWELMTFQAKPYDGIIPAREIDPLEKGERLPQPICTIDVYMI 944
 Qy 955 MKCKWMIDSECRPRFELVSESRMARDPQREYVIONEDLGASPLDSTFYKSLDEDDM 1014
 Db 945 ILCKWMIDSSRRPRFELVSESRMARDPQREYVIONEDLGASPLDSTFYKSLDEDDM 1000
 Qy 1015 GDLVDAEYLVPQGFPCPDPAAGAGVYHHRSSSTRSGGDTLLGLEPSEEBEAPRSP 1074
 Db 1001 -DVDADEYLLPYKRI--NRQGS-----EP 1022
 Qy 1075 LKPSSEGASDVFDGLGMAAKGLOSLPTHDSPLQRISEDTV-PLPSETDGYVAPLTC 1133
 Db 1023 CIPPTGH-----PVRENSITLRNISDPTQNALEKDLGH----- 1056
 Qy 1134 SFQPEYVNOPDVRPOP-----PSPRE-----GPLP-AAPRAGATLERAKTLSPGXGV 1180
 Db 1057 ----EYVNOPGETSETSRSLDIYNPNYEDLTDGMGVVSLSSCAENLFRPEYLNQNSL 1112
 Qy 1181 VKDVFARFGAVENPEYLTQGGAAPOHPFPAPFPAFDNLTYWDDPBERGAPSTFGT 1240
 Db 1113 ---PLVSSGSMDDPDY---QAG-----YQAF-----LPQTGALTNQWFL 1147
 Qy 1241 PTAENPEYLG 1250
 Db 1148 PAENLEYL 1157
 Db 1148 PAENLEYL 1157
 RESULT 9
 ERB3 HUMAN
 ID ERB3 HUMAN STANDARD; PRT; 1342 AA.
 AC P21660;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor (protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).
 GN ERB3 OR HER3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90083234; PubMed=2687875;
 RA Kraus M.H., Iseling W., Miki T., Popescu N.C., Aaronson S.A.;
 RT "Isolation and characterization of ERB3, a third member of the
 RT ERBB/epidermal growth factor receptor family: evidence for
 RT overexpression in a subset of human mammary tumors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=50311312; PubMed=2164210;

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
 RA Todaro G.J., Shoyab M.;
 RT "Molecular cloning and expression of an additional epidermal growth
 RT factor receptor-related gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
 RN [3]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=Placenta;
 RX MEDLINE=9328282; PubMed=7685162;
 RA Kato M., Yarak Y., Sugimura T., Terada M.;
 RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
 RT tyrosine kinase.";
 RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTRK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
 CC SECRETED (SHORT FORM).
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
 CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
 CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or
 CC send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; M29366; AAA35790.1; -;
 DR EMBL; M34309; AAA35979.1; -;
 DR EMBL; S61953; AAB26935.1; -;
 DR PIR; A36223; A36223.
 DR HSSP; P11362; IFK.
 DR Genew; HGNC:3431; ERBB3.
 DR MIM; 190151; -;
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; Fuf; 3.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
 DR PROSITE; PS00109; PROTEIN KINASE-TYR; FALSE NEG.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 19
 FT CHAIN 20 1342
 FT DOMAIN 20 643
 FT TRANSMEM 644 664
 FT DOMAIN 665 1342
 FT DOMAIN 709 966
 FT NP_BIND 715 723
 FT BINDING 742 742
 FT ACT_SITE 834 834
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT POTENTIAL.
 FT RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT EXTRACELLULAR (POTENTIAL).
 FT POTENTIAL.
 FT CYTOPLASMIC (POTENTIAL).
 FT PROTEIN KINASE.
 FT ATP (BY SIMILARITY).
 FT BINDING (BY SIMILARITY).
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 566 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 141 183 ELISGVYENKDKLMDTIDMDRIDVDRDAEIVYKONGR
SC -> GCFPMVSGLPQADQWYLLDDDDPRLLTSLASSK
VPVTLAAV (IN SHORT ISOFORM).
FT VARSPIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT SEQUENCE 1064 1064 E -> G (IN REF. 2).
SQ 1342 AA; 149097 MW; 7201E7F66CA374BD CRC64;

```

Query Match 35.1%; Score 2398.5; DB 1; Length 1342;
 Best Local Similarity 40.0%; Pred. No. 2e-119;
 Matches 527; Conservative 192; Mismatches 458; Indels 139; Gaps 33;

```

QY 10 GLLALLPPGAA--STQVCTGDMKRLPASPEPTHLMKRLHYQGGQVQGNLELYLEPT 67
DB 11 GLFLSLARGSEVNSQAVCPGTLNGLSVGDAMENOYQLYKLYERCEVVMGNLEIVLTS 70
QY 68 NASLSFLODIOEVQGVLLAHNOVRQVPLQRLRIYRGQTLFEDNALAVLNDGDDLTNT 127
DB 71 NADLSFLQIREVTGYLVANMFSTLPLNLRVARGTQVYDGKRAIFVM-----LNY 125
QY 128 PVTGASPGGLRELQRLSTLEILKGVLLIQRNPLCYODTILWKDIFHKNNQALTLIDTN 187
DB 126 ---NSSHALRQLRLTQLTEILSGVYIEKNDKLGMDTIDWRDIRD---AELVVD 178
QY 188 RSARCHRCSPMCKSGSKWESSEDOSLRRTYCAGGC-ARKGRLPTDCHGCAAGCTG 246
DB 179 NGRSCPCPHEVCCKG-RCMGPGSDECOITLTKTICAPQNGCHCGPMPNOCHECAGGSG 237
QY 247 PKHSDCLAFHFNHSGICELACPALVTYNTDFEESPNEBGRYTFGASCVTACPNVYST 306
DB 238 PQDTCFACRHNHNDGACVPRCPQPLVYMKLTFQLEPNHTYXQVGVAVASCPNFFV-V 296
QY 307 DVGSCVLVPLHNOETADGTQRCCKSKPCARVCYGLGMQYIANKSF--IGTLEH- 363
DB 297 DQTSQVRAQPPDKMEVD-RNGDKMCEPCGGLCPKACEGG-----SGSRFFQTVDSNDIG 350
QY 364 FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEOLOVFETLEETIGTVLSAPRDLPLDS 423
DB 351 FVNTCTILGNDLPFLITGLNGDFPMHKIPALDPEKLVFRVRIITGLYLIQSPRPMHNS 410
QY 424 VFQNTQVIRGRLIHNGAVS-LTLQGLISWIGLSRLRELSSALIHNTHLICFVHTVPM 482
DB 411 VFSNLTITIGRSILYNGFSLIMKNLNTVSLGFRSLIKETISAGRITISANRQLCYHNSLNM 470

```

```

QY 463 DQLEPNPQALLHTA-NREDECVGEGLAGHQLCARHGCKGPRPTQCVNGSQFLRGQECV 541
DB 471 TKVIRGPTEERLDIKHNRPRRDCAEGVVCPLCSSGGCWPGGQCLSCNRYSGVCV 530
QY 542 EECRVTLQSLPEVYNARHCLPCHPECOPONGSVTCFGEADQCVACAHYKDPFCVACP 601
DB 531 THCNFLNGEPPEFAHBAECFSGHPECQMEGTATONSGSDTCAQCAHFRGPHCVSCP 590
QY 602 SGVPRDLSYMEIMKFPDEGACOPCPINCTHSCVYLDLDDKGPAP-----QRASPLTS 654
DB 591 HGVAG--AKGIYKYPDVQNECRCHENCTGCKGPELDQCLGQTLVLIGKTHLTALTV 648
QY 655 FNNFTVSKWLAVPVKYSASHLEKROQKIRKTKTARLLQETLVEPLTPSGAMPQAOQRI 714
DB 649 IAGLVIFPM---LGGTFLYWRARRIQNKRAMRYLERGESIEPLPS-ERAKVULRI 703
QY 715 LKETELRKVLKLSGAFGTVYKGMIDGENVKIPVAIKVRENTSPKANEILDEAVYM 774
DB 704 FKETELRKVLKLSGVTGTVHKWVPEGESIKIPVCIKVIDSGSPQAVTDHMLAI 763
QY 775 AGVSPVYSRLIGICTSTVOLVTQMPYGLDHDVENRGLSGQDLNMCQIAKMS 834
DB 764 GSLDHAHIVRLIGCPGSSLOLVTVQYPLGSLDHDVHQRGALPQLLNQVIAKMY 823
QY 835 YLEDPVLRDLAARNVLVKS PNHVKTIDFGLARLDIDETEHADGKVPKIMMALESI 894
DB 824 YLEHGVHRLAARNVLVKS PSOVADPVALLPDDKQOLLYSEAKTIPKMALESI 883
QY 895 LRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICITDVYMI 954
DB 884 HFQKTHQSDVWSYGVTVWELMTFGAEYAGRLAELVADLLEKGERLAQPCITIDVYMV 943
QY 955 MYKCMIMDSERPPRELVSEFSMARDPQAFVYIQNEDLCPA---SPLDSTFYSRLIED 1011
DB 944 MYKCMIMDENIRPPEKELANEFTRMADPPRYLVIKRES-GGCIAPGPEEPGLTNKYLEE 1002
QY 1012 DDMGDLYDAEYVLPQCGFCPPDPAPAGWGHRRHSSSTRSGGDTLGLER-SEBEA 1070
DB 1003 VELERPELDLDLDAEED-----NLATTIGSALSLVGLNLRPG 1043
QY 1071 PRSPLABEGAGSDVFDGDLGMAKGLQSLPTHD-PSPLQRYSEDPVLP-----SE 1123
DB 1044 QSLLSPSSGY-MPMNQNLGSCQESAVSSSRCPVELH-----PMRGLASES 1096
QY 1124 TDGVYA-----PLTCSQPEP---YNOQPVNRQPSRPGP----- 1156
DB 1097 SEGHTGSEALQEKVSMCRSRSRSPRPGDSAYHSQRHSLTPTVPLSPGLDEEDV 1156
QY 1157 ---LPAAPAGATLERAKTLP--GRNGV-----KDVFAFGAVENPEYLTPOGGAAP 1205
DB 1157 NGYVMPDTHLKGTSRSGCTISSVGLSLVTEBEDD-----EVEYEMRRRRRSP 1208
QY 1206 QPAPPAPAFNDLYWD-----QDPPEGAPSTKGTPTANPEYL 1249
DB 1209 -PAPPSPSLBELGYEYWDVGSLSIGTQSCPLHPVPIPTAGTTPDEDEYEM 1263

```

RESULT 10
 ERB3 RAT STANDARD; PRT; 1339 AA.
 AC Q62799; Q62955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (C=erB3).
 GN ERB3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

Query Match	Similarity	Score	DB	Length
Query Match	34.0%	Score 2325.5;	DB 1;	Length 1339;
Best Local Similarity	40.5%	Pred. No. 1.5e-115;		
Matches 521;	Conservative 169;	Mismatches 436;	Indels 159;	Gaps 34
Sequence	1339 AA;	147545 MW;	0AA5F24028BFDPLE	CRG64;
FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	629		BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	333	353	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .)
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .)
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
Query Match	34.0%	Score 2325.5;	DB 1;	Length 1339;
Best Local Similarity	40.5%	Pred. No. 1.5e-115;		
Matches 521;	Conservative 169;	Mismatches 436;	Indels 159;	Gaps 34
Sequence	1339 AA;	147545 MW;	0AA5F24028BFDPLE	CRG64;
3	LALCPWGLLALLP	PGNA---	STOVCGTGTMKRLRLPAS	PETHLDMLEHLYOGCOVQGN
7	LQVLC---	FLSLARGSEMNSQAV	CGTINGLSTVGDA	NOQVTLKLYEKEVWNGN
60	LELTYP	LTNASLSFLQDI	QEVQGYVLI	IAHNOVROVLPQRLRI
63	LEIYLT	GHADLSELOW	IREVTGYLV	IAVMEFVSILPRLNLV
120	GDPLNT	PTVGA	BGRI	RELOLSLREILKGGVLI
121	---	INMYT---	NSSHALRQLK	PQVLEILSGGYV
180	ALTLID	TRSRACPC	SPMCGR	SKRCWBSSEDCSL
171	GAEIV	KKNGAN	CP	RPHEVCCKG-RQWGG
229	QQAAG	CTGPKASD	CLACI	HFHNSGICELHCA
230	ECAGG	CSPOPTD	CFACR	RFNDSGACVPRCE
239	CPYNYL	STDVSG	CTLVCP	ILHNOETVAEDG
290	CPHNFV	-VDQY	FCYRAC	PRDMEYD-KHGLK
357	IGITELE	FAGCKKIP	GSIALP	PBSFGDGPASNTA
348	ID---	GVNCTK	ILKILND	FLITLNVDPMKH
417	DSLPL	SLFOULQY	IRGRIL	HNAGVY--LTLOG
404	PHMNF	SVFSNLT	ITIGRSL	YNRGSFLILKMLN
476	FVHTV	PMWDOLE	FRNPHO	ALLHTA-NRPE
464	YHSL	ILNTRIL	LRGSEER	LIDIKYRPLGE
535	LRGDE	CEVECV	LVLOGL	PREVYNAHCLP
524	SREGV	CVTHCN	FLDGE	EPPEVFAHQ
595	FCVARR	CSGV	VPDL	SYWPIK

Db 584 HCVNCSCHGILG-AKPIYKTPDAQNERPCHEHNTQCCNGPELQDCIGQ-AVLWSK 639
 Qy 654 --I FNNFVSWLRVKKVSAHLEKRRQKIRKRYTRRLQETELVEPLTPSGAMPNOAQ 711
 Db 640 PLVIAVATGLAVLIMILGSGFLWGRRIQNKRAMRRLGEGESIEPLDPS-EKANKVL 698
 Qy 712 MLKLETELKRYKVLGSGAFVYVIGIWTPEDENKIPAIKVLBNKSPKANKETLDBA 771
 Db 699 ARIKRETELKRLKVLGSGVFTVHKGIWIPEGESIKIPICIVIEDKSGRQSFQAVTHM 758
 Qy 772 YMAVGASPVYRLLGICLTSTVQVLTQMLPGCLLDHVRNKGRLGSDLLNMCMOIAQ 831
 Db 759 LAVGSLDHAHIVRLILGCGSSLGQVLTQVYLPGLSLDHYKCHRELTLGQLLNMGVQAK 818
 Qy 832 GMSYLEDLVLRDLAARVLYKSPNHKITPFGALRLDIDETVHADGKVPKIMMAL 891
 Db 819 GMYLLEHSMVHRDLARNVLMKSPQVADFGVADLLPPDKOLLHSEAKTPKIMMAL 878
 Qy 892 ESILRRFTHSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPICTIDV 951
 Db 879 ESIHKGITHSDVMSYGVTVWELMTFGAEPIAGRLAEIPDLLEKGERLPQPICTIDV 938
 Qy 952 YMIWYKCMWIDSECRPRELVSFESRNARDQRFVJQNEMLGASPLDSTFYSLLED 1011
 Db 939 YMWYKCMWIDENIRPTFKELANFTRMARDPRYLVIKRAE-GRGTP-PAAPPSVLT 995
 Qy 1012 DDMGDLVDAEYLVPOQGFPCDPRAGAGWVHHHRSSSTSGGDLTLGLEBEE--- 1068
 Db 996 KEL-----QEALEPEL-----DLDLLEAEELGA 1021
 Qy 1069 -----EAPRPLAPSEG-----AGSDVFDGLGMAKAGQSLPT 1103
 Db 1022 TSLGSALELPTGLTRPRSGQLLSSSYMPMNGSLGECALDANVAGRGQFRPISL 1081
 Qy 1104 HPPSPLOKXSEPTVPLPSETGYV---APV-----TC-----SQPE---YVNP 1143
 Db 1082 H-FIPGR-----PASESGHYTGSSEALQEKVYCRSRSRSPRPRDSAYHQR 1133
 Qy 1144 DVRRPDPSPREP-----LPAARPAATLEPAKTSP-RKNGV-----KVVF 1185
 Db 1133 HSLTPVTPPLSPGLEEDNGVYVPTHTLRGASSSREGTSSVGLSSVLTGTEEDD- 1191
 Qy 1186 AFGAVENEYVLPQGAAPORHP 1210
 Db 1192 -----EEYEMMRKRGRSP-PRPP 1209
 RESULT 11
 EGFR_DROME STANDARD; PRT; 1426 AA.
 AC P0412; O61601; O9W2G0; P81868;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
 DE (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB).
 GN EGFR OR TOR OR C-ERBB OR DER OR CG10079.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NC NCB1; Taxid=7227;
 RN 11
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
 RX MEDLINE=94350209; PubMed=8070664;
 RA Clifford R., Schubach T.;
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
 RT subdomains of the receptor protein."
 RL Genetics 137:531-550(1994).
 RN (2)
 RP REVISIONS.
 RA Clifford R., Schubach T.;
 RN submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=85124611; PubMed=2982499;
 RA Linne E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
 RT "The Drosophila EGF receptor gene homolog: conservation of both
 RT hormone binding and kinase domains."
 RL Cell 40:599-607(1985).
 RN [4]
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
 RC STAIN=Oregon-R; Tissue=Embryo;
 RX MEDLINE=87002474; PubMed=3093080;
 RA Schneider E.D., Segal D., Glazer L., Shilo B.-Z.;
 RT "Alternative 5' exons and tissue-specific expression of the
 RT Drosophila EGF receptor homolog transcripts."
 RL Cell 46:1091-1101(1986).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=99102120; PubMed=9882502;
 RA Leeshkin A.M., Yu S.-Y., Katz J., Baker N.E.;
 RT "Several levels of EGF receptor signaling during photoreceptor
 RT specification in wild-type, Ellipse, and null mutant Drosophila."
 RL Dev. Biol. 205:129-144(1999).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
 RC STAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hopkins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borkov A., Botchan M.R., Bouck J., Brinkstein P., Brotilier P.,
 RA Butts K.C., Buzam D.A., Butler H., Cadiot E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spaulding A.C., Stepleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassenaar D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [7]
 RP SEQUENCE OF 959-1078 FROM N.A.
 RC STAIN=Daekwanryeong;
 RX MEDLINE=85137938; PubMed=2983232;
 RA Madsen S.C., Vincent W.S. III, Blodau-Wentworth D.;
 RT "A Drosophila genomic sequence with homology to human epidermal

Db	603	IADGYISNAAYK--FDNRTRCKICHECR-----TONGAGAHQCECVHVRGCGHCVSSC	654
Qy	601	P-----SGVY-----PDL	608
Db	655	PNKKYNDGVCRECHATCDGCTGPTDITIGACTTCNTAIINNDATVAKCLLKDKRCPD-	713
Qy	609	SYMPIMKF--DDEGACGP-----CPI-----NCTH-----	632
Db	714	GY--FMEEVHPQEGSLKPLAGRAVCRKHPLCELOTNNGVHEQYCSKTHHKREQGET	771
Qy	633	-----SC-----VDLDDKG-----	641
Db	772	ECPADHYDEORECFQRHPECNGCTGPGADCKCRNFKLPDAETGTVYNSTFNCTS	831
Qy	642	-CPAEOR-----ASPLTS--IFNN-----FTVFMIRPKVASHLEK--	676
Db	832	KCLEMRHVNYQTALIGPYCAASPSSKITANDVMNIFITGAVLPTIILCVYTI	891
Qy	677	-RROQKIKYT--MRLLQETELVPLTPSGMPCQAOMRIKETELRKVYLSGAGFT	733
Db	892	CRQKQKAKETVKMFMALSGCEDSEPLRPSNIGAMLCRLVKADELKKGGVLTGMGAGGR	951
Qy	734	VYKGIWIPDGENVKIPVAIKVLENTSPKANEIDEAVYVAGVSGPSVYSRLGCLFST	793
Db	952	VYKGVWVEGENVKIPVAIKELTKSTGASSESEFLREAYIMASEHVNLKLIAYCMSSQ	1011
Qy	794	VQLYTQQLMPYGLCLDHPVENRGRLGSSQDLLKWCQIAYKASYLEDVRLVHRDLAARNLY	853
Db	1012	MMLLITQMLPCLCLLDYVNNRDKIGSKALLMNSTQIAGMSYLEKRLVHRDLAARNLY	1071
Qy	854	KSPNHVKITDGLARLIDIDETEVYADGKVPDKMVALESLRRREFTHOSDPWASGYTV	913
Db	1072	QTPSLVKITDGLAKLLSSDSNEYQAAGSKPIKMLALECIRNRFTFSKSDVMWAGVYIM	1133
Qy	914	ELMTFGAKPYDGIPIAREIPDLLENGERLPORPICIITIDVYIMVWCMTDSECRPFRELY	973
Db	1132	ELLTFGQRPHNIPAKDIPDLIEVGLKECEPESGLDICTLLSCWHDAMRPFKQLT	1193
Qy	974	SEFSRMARDPQRFVVIQWEDLG--PASPLDSTFYRSLEDD--DMGSLVDAEVLVQO	1028
Db	1192	TVFMEFAPDPERYLAIIPDKFTRLPA-----YTSQEKDILKLAFTTQGESEIAPKD	1244
Qy	1029	GFECPPDPAFGAGVHHRRSSSTRSGGDLTLGLEPSEEAR-----RSPLAPSEAG	1082
Db	1245	DYLOPKAAFPS-----HRTDCT-----DEMFKLNRXCUPSNKNSGT	1283
Qy	1083	SDVFDG--DLGMGAKLQSLPHHDSPLQRYSDP--VPLPSEFDGVALTSGPOEY	1139
Db	1284	DDEDDSGAREVGVGNLR-----LDLPVEDDVLMP--TQCPGPN	1321
Qy	1140	VNCPDVBPQPSPEEGPLPAAPGATLERAKTLPGRNGVVKVDFVAFGAGVENEPEY--	1197
Db	1322	NNNNN-----NPNQNMMAAGVAAGYM-----DLIGPVSVDNPEYILN	1360
Qy	1198	--TFQGAAPQPH-----PPAFSP--ARDNLTYMD	1224
Db	1361	AQILGVSESPITPTQITIGVWGCGCTNEVKVPMFGSEPTSSDHEVYND	1408

ERBB_ALV	STANDARD;	PRT;	634 AA.
AC	P00534;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	15-JUL-1999 (Rel. 38, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112)		
GN	V-ERBB.		
OS	Avian leukosis virus.		
OC	Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.		
OX	NCBI_TaxID:11864;		
PN	[1]		

RP SEQUENCE FROM N.A.
RA MEDLINE=953286222. PubMed=2998784:
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RT Critchenden L.B., Raines M.A., Kung H.-U.;
RT "c-erbB activation in ALV-induced erythroleukosis: novel RNA
RT processing and promoter insertion result in expression of an
RL amino-truncated EGF receptor ";
RL Cell 41:719-726(1985).
CC
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).

CC	EMBL, M10066, AAA48763.1, ALT_INIT.
DR	PIR, A00643, TVCHLV.
DR	PIR, B00643, TVEVLV.
DR	HSSP, P11362, 1RGK.
DR	InterPro, IPR000719, Euk_Pkinase.
DR	InterPro, IPR01245, Tyr_Pkinase.
DR	Pfam, PF00069, Pkinase.1.
DR	PRINTS, PR00109, TYRKINASE.
DR	ProDom, PD000001, Euk_Pkinase.1.
DR	SMART, SMO0219, TyKc.1.
DR	PROSITE, PS00107, PROTEIN_KINASE_ATP.1.
DR	PROSITE, PS00109, PROTEIN_KINASE_TYR.1.
DR	PROSITE, PS50011, PROTEIN_KINASE_DOM.1.
KW	Transferase, Tyrosine-protein kinase; Oncogene; Glycoprotein, Phosphorylation.
FT	DOMAIN 132 399
FT	NP_BIND 138 146
FT	BINDING 165 165
FT	ACT_SITE 257
SC	SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

	Query Match	24.9%	Score 170.45	DB 1	Length 634
	Best Local Similarity	51.6%	Pred. NO. 4e-03		
	Matches	364	Conservative	72	Mismatches 149
				Indels 121	Gaps 15
QY	587	CAHYKDPCCVCAKCSGSKYKPDLSYPIKPFDEBGAQCPCEINCTHSQVLDLDDKCPAEQ	646		
Db	3	CAHFIIDGPHCVACACAGVATGENDTL-VKRYADANAVALCHPNCITRECKSGPLEGCE---	58		
QY	647	RASPLTSINNTFTVSEFMLRVPKVASHLKEROOKIRKYTRMRLLQETELVEPLTPSGAM	706		
Db	59	NGSKTSLIAGVAVGGILCLVYVGIGISGLYLRHHIVKRTLRLLQRELVPELTPSGEA	118		
QY	707	PNOAQRIILKETELAKRYKVLGSGAFGYTKQITLPDGEANYKIPAIIVLRENTSPKANKE	766		
Db	119	PNOAHRIILKETEFKRVKVLGSGAFGYTKQLMPEGEKXKIPAIKELREASTPKANKE	178		
QY	767	ILDEAYVAVGSPVYSRLGICLTSTVQLTQQLMPGCLLDHRENGRGLSGQDILNNC	826		
Db	179	ILDEAYVAVSVDPNPHVCRLLGICLTSTVQLTQQLMPGCLLDYIRHKONIGSQYILNNC	238		
QY	827	MOIAGMSVLEDPVRIYHBDLAARVAVLKSPHNYKITPFGIARLLIDIDETVYHADGKVP	886		
Db	239	VOIAGQNYLERRRLVHRFDLAARVAVLKTPOHVATITDFGLAKLLIGADEKXYHAEKGKVP	298		
QY	887	KWMALESILRRRFTHQSDVMSGYATVMEIIMPFGAKPVDGI-PAREIPDLLEKGERLLPQPI	946		

```

Db 299 KMALESILHRIYTHQDWSYGVTVWELMTFGSKPYDGIPIASEISVLEKGERLPQPI 358
Qy 947 CTIDVYIMVCKMIDSECRPRELVSEFSRMAADPQRFVIO-NEDLGASPLDSTFY 1005
Db 359 CTIDVYIMVCKMIDSDSRPKRELLAEFSKMARDPRLVIOGERHMLPSPTDSKFY 418
Qy 1006 RSLLEDDMDGLVDAEELVPOQGFPCDPAPAGGVHHRHSSSTRSGGDLTLGLEP 1065
Db 419 RLMEEEDMEDIVADEELVPHQGF-----NSPT----- 449
Qy 1066 SEEBAPRSP-----APSEGASDVDFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPL 1120
Db 450 -----SRPLSLSLASATSNNSANTCID-----RNGQGHPRVEDSFVQRYSSDPTGNF 496
Qy 1121 PSET--DGYVAPLTCSPQPEYVNOVDVPPSPRSGPLPAPAPAGATLEBAKTLSPGKN 1178
Db 497 LEESIDGFL-----PAPRYVNO-LMPKKPS-----TAMVON 527
Qy 1179 GVAVDVF-----AFGAIVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYM 1223
Db 528 QIYNNISLTAISKLPMDSRVONSHSTAVDNPEYL-----NTNOSPAAKTVEFESSPYW 579
Qy 1224 DQ-----DPE-----RGAPSTFGKGTPTAENPEYLGDPV 1254
Db 580 IQSGNHQIINLNDPQYQODFLPNETTKPGLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P0053;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=79685;
RN 11)
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=4026539; Pubmed=6333229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RL gene family.";
RL Cell 35:71-78(1983).
RN 12)
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=4223957; Pubmed=6328658;
RA Debure B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RT "Sequencing the erbb gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: K02006; AAA42394.1; ALT_INIT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A00644; TVRK.
DR HSPF: F11362; TVRK.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase_1.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; TyrKc_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR TRANSFERASE: Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBD0D6745D609 CRC64;

```

```

Query Match 24.3%; Score 1658; DB 1; Length 604;
Best Local Similarity 51.4%; Pred. No. 116-80;
Matches 354; Conservative 69; Mismatches 142; Indels 124; Gaps 14;

```

```

Qy 587 CAHYKDPFCVAPCGVCPDLSYMPFWKPPDEGACQPCPINCTHSCVDLDDGCCPMEQ 646
Db 3 CAHFDGPHCVYACPAAGVLTGENDTL-VKTIADANAACVCLCPNCTRGCKGGLGCP-- 58
Qy 647 RASPLTSIFNNFTVSFWLVPKVASHLKRRQCKIRKRYTRRLLOETELVEPLTPSGAM 706
Db 59 NGSKTPSIAGVGGGLCLLVVVGIGLYLRRRIIVKRTLRLLQGERELVEPLTPSGEA 118
Qy 707 PNOGMRILKTELKRYVVLGSGAFGVYKSIWIPDGENVKIPIAIXLRLNTSPKANKE 766
Db 119 PNOALRIKTEFEKXVVLGSGAFGITIKLWIPGEKXVIPIAIXLRLNATSPKANKE 178
Qy 767 ILDEAYVAVGSPVSRILGICTSTVQVLTQCLMPYGCILLDHYRENRGRSGODLWNC 826
Db 179 ILDEAYVAVSDNVHVCILLGICTSTVQVLTQCLMPYGCILLDHYRENRGRSGODLWNC 238
Qy 827 MOIAKMSYLDVLRHDLAARVLYKSPNHVKITDPGLARLLDIDETEVHADGKVP 886
Db 239 VOIAKMYVLEERLVRHDLAARVLYKTPQHVKITDPGLAKLLGADEKEVHAEGKVP 298
Qy 887 KMALESILRRRFTHOSDWSYGVTVWELMTFGSKPYDGIPIASEISVLEKGERLPQPI 946
Db 299 KMALESILHRIYTHQDWSYGVTVWELMTFGSKPYDGIPIASEISVLEKGERLPQPI 358
Qy 947 CTIDVYIMVCKMIDSECRPRELVSEFSRMAADPQRFVIO-NEDLGASPLDSTFY 1005
Db 359 CTIDVYIMVCKMIDSDSRPKRELLAEFSKMARDPRLVIOGERHMLPSPTDSKFY 418
Qy 1006 RSLLEDDMDGLVDAEELVPOQGFPCDPAPAGGVHHRHSSSTRSGGDLTLGLEP 1065
Db 419 RLMEEEDMEDIVADEELVPHQGF-----NSPT----- 449
Qy 1066 SEEBAPRSP-----APSEGASDVDFDGLGMAAGLQSLPHDPSPLQRYSEDPTVPL 1120
Db 450 -----SRPLSLSLASATSNNSANTCID-----RNGQGHPRVEDSFVQRYSSDPTGNF 496
Qy 1121 PSET--DGYVAPLTCSPQPEYVNOVDVPPSPRSGPLPAPAPAGATLEBAKTLSPGKN 1178
Db 497 LEESIDGFL-----PAPRYVNO-LMPKKPSYAM----- 524
Qy 1179 GVAVDVF-----AFGAIVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYM 1223
Db 528 QIYNNISLTAISKLPMDSRVONSHSTAVDNPEYL-----NTNOSPAAKTVEFESSPYW 575
Qy 1220 LYVWDQDPPERGAPPSFFKGTPTAENPEY 1248

```

Db 576 SPYWIGSGNHQ-----INLDNPDY 594

RESULT 14

ERBB_AVIEW

ID ERBB_AVIEW STANDARD; PRT; 540 AA.

AC P11273;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).

GN V-ERBB.

OS Avian erythroblastosis virus (strain ts167).

OC Viruses; Retroviridae; Avian type C retroviruses.

OX NCBI_TaxID=103998;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87064458; PubMed=2878364;

RA Choi O.R., Trainor C., Graf T., Beng H., Engel J.D.;

RT "A single amino acid substitution in v-erbB confers a thermostable phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells."

RT Mol. Cell. Biol. 6:1751-1759(1986).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib.ch).

CC EMBL: M13179; AAA2401.1; -

DR PIR: A25231; TYFVB.

DR HSSP: P1362; IFK.

DR InterPro: IPR000719; Euk_pkinase.

DR InterPro: IPR001245; Tyr_pkinase.

DR Pfam: PF00069; pkinase; 1.

DR ProDom: PD000001; Euk_pkinase; 1.

DR SMART: SM00219; TyrKc; 1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.

DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.

DR Transferrase: Tyrosine-protein kinase; ATP-binding; Oncogene;

KW Glycoprotein; Phosphorylation.

FT DOMAIN 132 399

FT NP BIND 138 146

FT BINDING 165 165

FT ACT_SITE 257 257

FT VARIANT 270 270 H -> D (IN THERMOSTABLE V-ERBB).

FT SEQUENCE 540 AA; 60412 MW; 5853297AA06865D CRC64;

Query Match 23.2%; Score 1585; DB 1; Length 540;

Best Local Similarity 54.2%; Pred. No. 6.8e-77;

Matches 335; Conservative 62; Mismatches 133; Indels 88; Gaps 12;

587 CAHYDPPFCVAVRCGVPDLVYPIKPFDEGACOPCPINCHSCVDDDDGCPAPQ 646

Db 3 CAHFDGPHCVACAGVAGLGEVDLT-VKRYADANAVCLCPNCRGKGGLESCP-- 58

QY 647 RASPLTSTFNNFTVSFMIRVPKVSASHLEKRRQKIRYVNRLLQETELVEPTPSGAM 706

Db 59 NGSKTPSIAGVVGGLLVVVGIGLGLYLRHRHVRKRTLRLLOERELVEPTPSGEA 118

QY 707 PNOAMRIKETELRKVKVLSGAGAGTYKGIWIPDGENVYIPVAKYLRRENTSPKAKE 766

Db 119 PNOAHLRIKETEFKVKVVLGFGAGTYKGMIPGSKVITPVAIKELREATSPKAKE 178

QY 767 ILDEAYVAGVGSFVSRLGLCLTSTVQLVTLQMLPYGCLLDHVENNGRLGSDLLNMC 826

Db 179 ILDEAYVAGVGSFVSRLGLCLTSTVQLVTLQMLPYGCLLDHVENNGRLGSDLLNMC 238

QY 827 MOIAKMSYLEDEVALVHRDLAARNVLYKSPNHWKITDGLARLUDIDETEHADGKVP 886

Db 239 VOIAKMNVLTERHMHVHDLAARNVLYKTPQHVKITDGLAKQGADEKEYHAGKVP 298

QY 887 KMMLESLRRRFLTHQSDVWSYGVTELMTFGAKPVDGIPARITPDLLEGERLPQPI 946

Db 299 KMMLESLHRIYHSDVWSYGVTELMTFGSKPYDGI PASISVLEGERLPQPI 358

QY 947 CTIVVMIMVCMIMIDECRPFRELSEFSRMRDPQRFVIO-NEDLGASPLDSFTY 1005

Db 359 CTIVVMIMVCMIMIDECRPFRELSEFSRMRDPQRFVIO-NEDLGASPLDSFTY 418

QY 1006 RSLLEDDMGDLVDAEYLVPQGFCEPDPAAGAGVHHRSSSTRSGGDLTLGLEP 1065

Db 419 RSLLEDDMGDLVDAEYLVPQGFCEPDPAAGAGVHHRSSSTRSGGDLTLGLEP 449

QY 1066 SEEDAPSPPL-----APSEGGSDVPDGDGMGAQKLGSLPTHPDPSPLQKYSDDPTVPL 1120

Db 450 -----SRTPLSLSLSTSNNSATNCIDRNG-----H----- 476

QY 1121 PSETDGYVAPLTCSPQEVYVQPDVRFQPSFREGPPAARPAAGAT-LEPAKTUSPGNG 1179

Db 477 PVREDGFL-----PADEVYVQ--LMPKXSTAVVQVQIVYISLTSKLPIDSRVQN- 527

QY 1180 VKVDFPAGCAVNEPFL 1197

Db 528 -----SHSTAVDNEPFL 539

RESULT 15

EGFR_CHICK

ID EGFR_CHICK STANDARD; PRT; 703 AA.

AC P13387;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER Fragment).

GN EGFR.

OS Gallus gallus (chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OX NCBI_TaxID=9931;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88261272; PubMed=3260329;

RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M., Ullrich A., Vennstrom B., Schlessinger J., Givol D.;

RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mouse cells, and differential binding of EGF and transforming growth factor alpha."

RT Mol. Cell. Biol. 8:1970-1978(1988).

RT FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF. AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to dimerization, internalization of the EGF-receptor complex, induction of the tyrosine kinase activity, stimulation of cell DNA synthesis, and cell proliferation.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib.ch).

or send an email to license@sb-sib.ch).

CC -----
 EMBL: M20386; AAA48760.1; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; Fu; 4.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; PARTIAL.
 DR PROSITE: PS00109; PROTEIN_KINASE_TIR; PARTIAL.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; PARTIAL.
 DR Transmembrane; Glycoprotein; Receptor; Signal; transferase;
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 30
 FT CHAIN 1 >703
 FT DOMAIN 31 654
 FT TRANSMEM 655 667
 FT DOMAIN 668 >703
 FT DISULFID 197 206
 FT DISULFID 201 214
 FT DISULFID 222 230
 FT DISULFID 226 238
 FT DISULFID 239 247
 FT DISULFID 243 255
 FT DISULFID 258 267
 FT DISULFID 271 298
 FT DISULFID 302 314
 FT DISULFID 318 333
 FT DISULFID 336 340
 FT DISULFID 513 522
 FT DISULFID 517 530
 FT DISULFID 533 542
 FT DISULFID 546 562
 FT DISULFID 565 581
 FT DISULFID 569 589
 FT DISULFID 592 601
 FT DISULFID 605 627
 FT DISULFID 630 638
 FT DISULFID 634 646
 FT CARBOHYD 134 134
 FT CARBOHYD 190 190
 FT CARBOHYD 200 200
 FT CARBOHYD 359 359
 FT CARBOHYD 368 368
 FT CARBOHYD 420 420
 FT CARBOHYD 573 573
 FT CARBOHYD 578 578
 FT CARBOHYD 613 613
 FT CARBOHYD 633 633
 FT CARBOHYD 648 648
 FT NON_TER 703
 SQ SEQUENCE 703 AA; 77427 MW; AAF2DE11B735A690 CRC64;

Query Match 22.7%; Score 1550; DB 1; Length 703;
 Best Local Similarity 43.8%; Pred. No. 6,4e-75;
 Matches 310; Conservative 104; Mismatches 265; Indels 28; Gaps 12;

QY 8 RGLLLALPPGAA-----STQVCTGDMKRLPLASPEETHLDMRLHYQGCQVQGNLE 61
 DB 13 RGAAYLVLLLLGVALCSAVEKVCQGTNNKLTOLGHVEDHFTSLQPMYNNCEVYLSNLE 72
 QY 62 LTLVPTNASLSFQDIQEVGVYLAHQVQVPLQRLIVRGSTOLFEDRYALAVDNGD 121
 DB 73 ILYVENHRDITFLKTIQEVGVYLAHQVQVPLQRLIVRGSTOLFEDRYALAVDNGD 132
 QY 122 PLNNTTPTVGTASPGGLRETELQSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNQAL 181
 DB 133 -NNKTO-----GLRELPKRLSELINGVKSINNPKLCNMDTVLWMDIIDTSRK-PL 182
 QY 182 TLID-TVRSRACHPCSPMGSGRCWGSSEDCSLTRTVAGGCA--RCKGPLPTDCHEO 239

DB 183 TLDPFASNLSSCPKPCPNCTEDHGWGAGEQNCQTLTKVICAQCCSGRCRGHPSPDCCNQ 242
 QY 240 CAAGCTGPKHSDCLACHFNHSGICEHLPLVYNTDTEFBSMNPREGRYTFGASCYTAC 299
 DB 243 CAAGCTGPRESDCLACRFDDATCTKTCPLVLYNPTTYMDVNPBEKISFGATCAREC 302
 QY 300 PNYLSTIDVSGCTLVCPPLHNOEYVAEDGTORCEKSKPCAVCYGLAQVYI KANSKFIGI 359
 DB 303 PHNYVTVDHSGCVSRSCNTDYEV--EENGVRCCKKCDGLSKVNCNGIGIGELKGLIS -INA 360
 QY 360 TELE-FACCKIIFGSLAFLESPDGPASNTAPLOPELOVFELEITGLIYISAWPDS 418
 DB 361 TINSFXNCTKINGDVSTILPVAFLGDAPTKTLPDPKTLDFVFRVKEISGFLLLQAMPDN 420
 QY 419 LPDISVFNQIVIRGRIIHNAGAVSLTQGLISWIGRLSRLASGLALIHNNFTLCPVH 478
 DB 421 ATDLYAFENLEIRRTQHOQYSLAVVNLQISLGRSLKEISGDLAIKKNKTLCTAD 480
 QY 479 TVPMDOLFRRPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPPTQCVNCSOFLRGQ 538
 DB 481 TNNRSLPATQSQKTKIIONENKNDCTADRHVCDPLCSDVGCWGPGRPHCFSPFSRQK 540
 QY 539 ECVRECVTLQGLPREYVNAARHCLPCHPECPONG---SVTQGPADQCVACAHKDPF 595
 DB 541 ECVKCNILQGBPRFERDSKCLPCHSECLVQNSTAYNTTSGGEPDHCKCAHFDGPH 600
 QY 596 CVARCPSGVKEDLSYMPIWKEPDEBGAQCPPICTHSCVDLDDKGCFAEQRASPLTSIF 655
 DB 601 CYKACPAGVLENDTL-VMKYADANAVALQLCHPNOTRCCKPGEGCP---NGSKTPSIA 656
 QY 656 NNFTYSFWLRPKYSASHLEKRRQOKIKYTMERLDETLEVEPLTP 702
 DB 657 AGVGGLLCLVVGVLGIGLYLRRIYAKRILRLRLORREIVEPLTP 703

Search completed: July 22, 2003, 03:18:05
 Job time : 21.642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.9246 Seconds

(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-632-652-14

Perfect score: 6834
Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTANPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriophage: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6125	89.6	1259	6	O18735
2	3107	45.5	1209	1	O9QX70
3	3078	45.0	1210	11	O9EP98
4	2685	39.3	1165	13	O9YH0
5	2646.5	38.7	1137	13	O9W674
6	2260	33.1	1328	13	P79754
7	1988.5	29.1	1433	5	O9B1H9
8	1871	27.4	419	4	O9UK79
9	1739	25.4	367	11	O8R2X1
10	1697.5	24.8	412	4	O8WYV0
11	1675	24.5	729	15	O86712
12	1673	24.5	567	15	O86714
13	1608.5	23.5	962	15	O64895
14	1600	23.4	545	15	O85468
15	1506.5	22.0	655	11	O9WVFS
16	1490.5	21.8	643	11	O9ERV6

17	1247	18.2	1193	5	O9Y1X8	O9Y1X8 ephydaria f
18	1181.5	17.3	1368	5	O23821	O23821 caenorhabdi
19	1150	16.8	1717	5	O26566	O26566 schistosoma
20	1126	16.5	527	13	O90836	O90836 gallus gall
21	1001.5	14.7	478	11	O9E5E0	O9E5E0 rattus norv
22	942.5	13.8	599	13	O9B9H2	O9B9H2 gallus gall
23	906	13.3	165	4	O14256	O14256 homo sapien
24	887	13.0	176	11	O923V5	O923V5 rattus norv
25	806.5	11.8	346	13	O8S2W1	O8S2W1 xiphophorus
26	778	11.4	435	5	O8S2W1	O8S2W1 xiphophorus
27	754.5	11.0	311	13	O99162	O99162 xiphophorus
28	739.5	10.8	1362	13	O9PV24	O9PV24 xenopus lae
29	734	10.7	331	4	O9BUD7	O9BUD7 homo sapien
30	723	10.6	149	6	O9B6G6	O9B6G6 oryctolagus
31	715	10.5	1671	5	O9NVV5	O9NVV5 biophthalari
32	688	10.1	1418	13	O93457	O93457 scophthalmu
33	675.5	9.9	1368	13	O8UW85	O8UW85 paraliichthy
34	673.5	9.9	1369	13	O8UW86	O8UW86 paraliichthy
35	657.5	9.6	1472	5	O9U5A8	O9U5A8 bombyx mori
36	657	9.6	1358	13	O73798	O73798 xenopus lae
37	652	9.5	1412	13	O8UW84	O8UW84 paraliichthy
38	635.5	9.3	1418	13	O8UW83	O8UW83 paraliichthy
39	631	9.2	1245	13	O9YGH8	O9YGH8 scophthalmu
40	629	9.2	1371	11	O9OYV4	O9OYV4 rattus sp.
41	621.5	9.1	2144	5	O9VDS4	O9VDS4 drosophila
42	589.5	8.6	1091	4	O9UMQ4	O9UMQ4 homo sapien
43	587.5	8.6	1036	4	O07912	O07912 homo sapien
44	586.5	8.6	1055	11	O54967	O54967 mus musculu
45	580.5	8.5	1072	4	O9BTB0	O9BTB0 homo sapien

ALIGNMENTS

RESULT 1
ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Erib-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
CX NCBI_TaxID=9615;
KM [1]
RN SEQUENCE FROM N.A.
RP YOKOTA H.;
RA "CDNA cloning of erib-2 from canine mammary gland."
RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSBP; P11362; IFGK.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Btk_kinase.
DR InterPro: IPR002174; Furin_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Btk_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137969 MW; E3364D4C4ACD46 CRC64;

Query Match 89.6%; Score 6125; DB 6; Length 1259;
 Best Local Similarity 89.8%; Pred. No. 0;
 Matches 1132; Conservative 40; Mismatches 92; Indels 6; Gaps 2;

QY 1 MEALALCRWGLLALIPPAASTOVCCTGTDMKRLPASPETHLDMRLHYGCCQVYQNL 60
 DB 1 MEALAMCRWGLLALIPSGAAGTQVCTGDMKRLPASPETHLDMRLHYGCCQVYQNL 60
 QY 61 ELTYLPTNASLSFLQDIOEVQGVYLAHQVQVPLQRIYRGTQLPFDNVALVLDNG 120
 DB 61 ELTYLPANASLSFLQDIOEVQGVYLAHQVQVPLQRIYRGTQLPFDNVALVLDNG 120
 QY 121 DPLNNTPTVGASPGGLRELOLRSLTEILKGVYLIQRNPQCYDTIWKDI FKHANOLA 180
 DB 121 DPLEGGIPAPGAAGGLRELOLRSLTEILKGVYLIQRNPQCYDTIWKDI FKHANOLA 180
 QY 181 LTLITNRSRACHPGSPMCKSGSCWSESDCQSLTRTACAGCARCKPRLTDCHEQC 240
 DB 181 LTLITNRSRACHPGSPMCKSGSCWSESDCQSLTRTACAGCARCKPRLTDCHEQC 240
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDPESMNPEGRTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDPESMNPEGRTFGASCVTACP 300
 QY 301 YNLSTDVSGCTLVCEFLHNOEYTAEDGTQRCCKSPCARVCYGLMOYIKANSKFIGIT 360
 DB 301 YNLSTDVSGCTLVCEFLHNOEYTAEDGTQRCCKSPCARVCYGLMOYIKANSKFIGIT 360
 QY 361 ELEFACCKIFGSLAFLPESFDGDPASNTAPLOEQVFEITETGYLYSAMPDILP 420
 DB 361 IOEFACCKIFGSLAFLPESFDGDPASNTAPLOEQVFEITETGYLYSAMPDILP 420
 QY 421 DLSVFONLOYIRGRIIHNGAYSLTLOGGISWLGRLSLELGSGLAIHHNTHLCRVHTV 480
 DB 421 NLSVFONLOYIRGRIIHNGAYSLTLOGGISWLGRLSLELGSGLAIHHNTHLCRVHTV 480
 QY 481 PMDOLFRNPHOALHTANRPEDECEVGLACHQLCRGHCWGPQPCVNGSOPFRGOC 540
 DB 481 PMDOLFRNPHOALHTANRPEDECEVGLACHQLCRGHCWGPQPCVNGSOPFRGOC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNSVTCFGEADQVACAAYKDPFCVARC 600
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECPQNSVTCFGEADQVACAAYKDPFCVARC 600
 QY 601 PSGVRPDLSPMPKWKPEDEEGACQPCINCTHSCVLDKCGPARASPLSTINFTV 660
 DB 601 PSGVRPDLSPMPKWKPEDEEGACQPCINCTHSCVLDKCGPARASPLSTINFTV 660
 QY 661 SEWLRVPKVSASHLEKROOKIRKYMTRLLOETELVEBLTSGAMPNQAORILKETEL 720
 DB 661 ILLAVVGLVLGILIKRRQKIRKYMTRLLOETELVEBLTSGAMPNQAORILKETEL 720
 QY 721 RKXKVLGSGAFGVYGIWIPDGENKIPVAIKVLENTSPKANKIIDEAVYMAVGSP 780
 DB 721 RKXKVLGSGAFGVYGIWIPDGENKIPVAIKVLENTSPKANKIIDEAVYMAVGSP 780
 QY 841 LVHRDLAARNVLYKSNHVKITDFGLARLLDIDETRYADGKVPKMALESILRRFT 900
 DB 841 LVHRDLAARNVLYKSNHVKITDFGLARLLDIDETRYADGKVPKMALESILRRFT 900
 QY 901 HOSDVMYSGVTWELMTFGAKPYDGI PAEIPDLLEKGRILQOPICITDVMYMKCM 960
 DB 901 HOSDVMYSGVTWELMTFGAKPYDGI PAEIPDLLEKGRILQOPICITDVMYMKCM 960
 QY 961 IJSECRPRRELVSFSSRNAPDORFVIONEDLGPASDLSTFYRSLDEDDMDGLVDA 1020
 DB 961 IJSECRPRRELVSFSSRNAPDORFVIONEDLGPASDLSTFYRSLDEDDMDGLVDA 1020
 QY 1021 EEVLVPQGGFPCDPAPAGAGVNHHRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080

DB 1020 EEVLVPQGGFPCDPAPAGAGVNHHRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1079
 QY 1081 AGSDVFPDGLGMAAKGLQSLPTHDPPLQRYSDPVPPLSEFTDGVYAPLTCSPQREYV 1140
 DB 1080 AGSDVFPDGLGMAAKGLQSLPTHDPPLQRYSDPVPPLSEFTDGVYAPLTCSPQREYV 1139
 QY 1141 NQPDVPPQPSPPRGPVPAAPAGATLER-----AKTLSPGKNGVNDVFAFGAVENPE 1195
 DB 1140 NQPDVPPQPSPPRGPVPAAPAGATLER-----AKTLSPGKNGVNDVFAFGAVENPE 1195
 QY 1196 YLTPQGAAPQHPHPAPSPAFDNLVYWDQDPPERKAPSPSTFKCTPTAENPEYGLDVPV 1255
 DB 1200 YLTPQGAAPQHPHPAPSPAFDNLVYWDQDPPERKAPSPSTFKCTPTAENPEYGLDVPV 1259

RESULT 2
 Q9QX70 PRELIMINARY; PRT; 1209 AA.
 AC Q9QX70;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=9025888; PubMed=2342466;
 RA Pech L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;
 RA Earg H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 encoded by an alternatively spliced transcript in normal rat tissue."
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Pech L.A.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K.; Dawson T.L.; Earg H.S.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 DR EMBL; M37294; AAF1408.1; -
 DR HSSP; P11362; IFCK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 965EE7F6C1B773 CRC64;

Query Match 45.5%; Score 3107; DB 11; Length 1209;
 Best Local Similarity 49.8%; Pred. No. 1.8e-228;
 Matches 636; Conservative 161; Mismatches 366; Indels 114; Gaps 26;

QY 3 LALCRWGLLALIPPA-ASTOVCCTGTDMKRLPASPETHLDMRLHYGCCQVYQNL 61

Db 15 LAALCAAG-----GALBEKKVCGGTSNRLTQGTFFEDHFLSLORMFNNECVLGNLE 66
 Qy 62 LTYLPTNASTLSFODIQEVOGYLLAHNOVOPVLRRLIVRGTLQEDFNVALVINDG 121
 Db 67 IYVORNDYLSFKTIQEVAGYVLLALNVERPLENLOIIRGNALYENTVALAVLSN-- 124
 Qy 122 PLNNTPVYASPGSLRELOLRLTEILKGVILQBNPOLCQODTILMDIHKHONQAL 161
 Db 125 -----YGTNKTGIRELPMRLQELLGAVRFSNNPLILCWNELIOMRDIY-QDVFLSN 175
 Qy 182 TLIDTNRS-RACHPCSPMKGSFRCWSESSEDCQSLTRVYACAGCA-RCKGRLPTDCHEQ 239
 Db 176 MSMDVQRHLTGCPKCDPCPCNCGSCMGCEENCCKLTIKICAOQCSCRRGGRSPDCCHNQ 235
 Qy 240 CAAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDFFESMNPNEGTYTGASCYTRAC 299
 Db 236 CAAGCTGPPESDCLVCHRFDEATCKDTCPMLNLPPTYQDVNVEGYSFGAATCVKXC 295
 Qy 300 PNYVLTSDVGSCTLVCPLNQEVTAEDGTQRCCKSPCARCYGLGWCYIKANSKFISI 359
 Db 296 PRVYVLTDHGSCVRACGPDYEV-EEDGVSKCKCKCGPRKVCNGIGIEFR-DTLSTNA 353
 Qy 360 TELE-FAGCKITFGSLAFPESEGDPPASNTAPLOEQLOVEETLEITGYVYISAMPDS 418
 Db 354 TIKKFKYCTAISGDHLIPVAFKGSFTRTPPLDPRELEILIKYVEITGFLILQAMBN 413
 Qy 419 LPLSYFONLOVYRGRILHNGAVSLTGLGISMILRELESGGLLHNNHLCFPH 478
 Db 414 WTDLHAFELREILIRGTRKHGQPSLAVGLNLTSLGRLKXISODVYIISGNRLCYAN 473
 Qy 479 TVPMDOLFENPHQALHTANREDECEVGEGLACHQICAGHCGWGPQTQCVNCSQFLRQ 538
 Db 474 TINKKLFSTPQKTKIMNRAKCKOKATNHVCNPLCSGCGWGPETDVCSCQVNSGR 533
 Qy 539 ECVBERVLOGLPREVYVNRHCLPCHRECPONGSVTCGCPADOCVACAHKDPFCA 598
 Db 534 ECVDCNILEGEPREVENSECTIQHPECLPTQNTITCGRPDMCICAHVDDPHCVK 593
 Qy 599 RCPGSKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKCAEORASP-LTSFNN 657
 Db 594 TPCSGIMGENNTL-VMKFADANNVCHLCHANCTYGCAGBGLKGC--QOPEGKISINTG 650
 Qy 658 FVTSFWLKYKVSASHLEKROOKTRKTYMRLLLOETLVEPLTSGAMPQONRIKE 717
 Db 651 IVGGLLFTVVALGIGLFMRRLQVRLKRLLOERELVEPLTSGAPNQHRLIKE 710
 Qy 718 TELRKVYLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANEILLDEAYVMASV 777
 Db 711 TEFKIKYLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANEILLDEAYVMASV 770
 Qy 778 GSPVYSRLIGICLTSTVOLVTLQMPYGLLDHVRNRLSGODLLNMCQIAKMSYLE 837
 Db 771 DNPHVCRLLIGICLTSTVOLVTLQMPYGLLDHVRNRLSGODLLNMCQIAKMSYLE 830
 Qy 838 DVLVLRDLAARNVYKSPBNHYKIDPGLARLLDDEHYADAGKVIKMMALESTLR 897
 Db 831 DRLVLRDLAARNVYKSPBNHYKIDPGLARLLDDEHYADAGKVIKMMALESTLR 890
 Qy 898 RFTHOSDVMSYGVIVWELMTGAKFYPDIPAREIPDLLEKSGRLPQPICTIDYVMWV 957
 Db 891 IYTHOSDVMSYGVIVWELMTGAKFYPDIPAREIPDLLEKSGRLPQPICTIDYVMWV 950
 Qy 958 CWMIDSEGRPRELVEFSKMARPOFVVIQ-NEDIGPASPDLSTFYRSLLEDDMGD 1016
 Db 951 CWMIDSEGRPRELVEFSKMARPOFVVIQ-NEDIGPASPDLSTFYRSLLEDDMGD 1010
 Qy 1017 LVDAEYLVPQGFCEPDPAFAGAGVMHRRSSSTRGCGDILTGLEPSEENPRSLA 1076
 Db 1011 VVDADEVYLIPQGF-----NSPT-----SRPL 1036
 Qy 1077 PSEAGSDVFPDGLCMGAKLQSLFTHDPSLPQRYSDPTVPPSEH--DGYVAPLTC 1134
 Db 1037 SSLSANSN-----SSTVACINNGSCRVKEDAFLORYSSDSPVSLTEDIIDDTFL----- 1086

Qy 1135 PPEVYNQPDVPPPPPPRECPPLPAPAGATLERAKTLSPCKGKGVYKDVAFAGVAVNP 1194
 Db 1087 PVPEYING-SVPRKPAQSVQPNVYHNCILHP-----ABGRDLHYQN--PHSAVANP 1135
 Qy 1195 EYL-TPQGAAPQHPHPAPFPAPFDNLVYMQ-----DP-----TERGARPPSTF 1237
 Db 1136 EYLNTAQ-----FTCLSSGFPSSALWIKGSHQMSLDNPYQODFFPEAKPNCIF 1186
 Qy 1238 KGTFAENPEYGLDVP 1254
 Db 1187 KG-PTAENAEYLRVAP 1202
 RESULT 3
 Q9EP98
 AC Q9EP98 PRELIMINARY, PRT, 1210 AA.
 DT 01-MAR-2001 (TREMREL. 16, Created)
 DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
 DT 01-JUN-2002 (TREMREL. 21, Last annotation update)
 DE Epidermal growth factor receptor isoform 1.
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVAC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maile N.J.,
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Scheil C., Pearall R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maile N.J.,
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275366; AAG28045.1; -
 DR EMBL; AF275364; AAG28045.1; JOINED.
 DR EMBL; AF275365; AAG28045.1; JOINED.
 DR EMBL; AF275367; AAG24386.1; -
 DR HSP, P11362; IFGK.
 DR MGD; MGI:95294; Egfr.
 DR InterPro: IPR000345; CytC heme bind.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-1like.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-1like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recept_L domain; 2.
 DR PRINTS; PR00109; TRYKINASE.
 DR PRODOM; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKc; 1.
 DR SMART; SM00219; TYKc; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA; 134840 MM; 62CD021C9DE32E18 CRC64;

Query Match 45.0%; Score 3078; DB 11; Length 1210;
Best Local Similarity 49.3%; Pred. No. 3e-226;
Matches 627; Conservative 163; Mismatches 372; Indels 110; Gaps 23;

```

11 LLLALPPGAA--STVCTGTDMLKLPASPETHLDMLRLHLYGQGVVQGNLELYLPTN 68
14 LLLALCAAGALEKKVCCQTSNRLTGLTFEDHFLSLGRMYNNCEVVLGNLEIYYQGN 73
69 ASLSFLQDIQEVQGYVLIHANQVYPLQRLAIYRGTOUPEENYALAVLDNDGPLNNTTP 128
74 YDLSFLKTIQEVAGVYLIANTVERIPLNLQIRSNALYENTYALALISN----- 124
129 VTGASPGIRELEQSLSTELKGVLIQGNPQCYQDTILMKDI---FHKNQDLATLI 184
125 -YQNRITGLRELPMRLQGLIGAVFNSNPILCNMDITQMRDIYQNFMSMSMDL---- 180
185 DTRSRACHPCSPCKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGK 243
181 -GSHPSCKPCDPSCKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGKSGK 239
244 CTGPKHSDCLACIENHSGICELCPALTYNTDPESSMPNPEGRYFGASGVTAQPKY 303
240 CTGPRSDCLVCKFQDEATCKDTCPPLMLNPTTYQMDVNEBKISFGATCVKCKPKRY 299
304 LSTDVSGCTLVCELNQEVYTAEDGTQCEKSKPCARVCGLGMOYIKANSKFIGITBLE 363
300 VVTDHSGCYRACGPDYEV--EEDGIRKCKKCKGCPCKVNGIGIGIFK-DLISINATIK 357
364 -FAGCKKIFGSLAFIPESFPDGPASNTAFOEQVFTLEITGILYLSMWPSLDL 422
358 HFXYCTAISGDHLILVAFKGSFTTPPLDPELEILTVETIGFLIQMPNNWDL 417
423 SVFQNTQVIRGRILHNGAVSLTLQGLISMLGRSLREKSGSLALIHNTHLCFHYTPW 482
418 HAFENLEIRGRIRKQGGFSLAVVGNLISLGRSKESDDVLIISGRNLCYANTINW 477
483 DQLEFNPHQALHTANRPDECEVSGELACHQICARHCHGPGPTQVNCSTLRQCEVE 542
478 KKLFGTPNQKTKIMNNRAKDCAVNVNCPICSSGCGPPEPRDVCQNSRRECEVE 537
543 ECEVLQGLPREVYNAHCLPCHPECPQNGSVTCFEPADQCVACAHYDPPFCVAPCS 602
538 KNIILGEPREFEINSECTQCHPECLPQAMNITCTGRGDNLCIQAHYDGHCVKTCDA 597
603 GVKPDLSTYMPWKPDEGACQPCPINCSTSCVDLDDKCPABQRAPLTS:FNFTVSF 662
598 GINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQCEVWPSGPKIPSIATGIVGL 656
663 WLRRPKVASHLEKRGQOKIRKYTMRLIOETELVEPLPPSGAMPNQAARILKTELK 722
657 LFLVVALVGLFMRRAHVRKRLRLQERLEVEPLPSEAPQAALRLIKTEFEK 716
723 VKVLGSGAGTYVKGIMIPDGENVKIPVAIKVLENTSPKAKELIDEAIVVAGYSPV 782
717 IKVLGSGAGTYVKGIMIPDGENVKIPVAIKVLENTSPKAKELIDEAIVVAGYSPV 776
783 SRLIGICTSTVQLVQLMPYGLDHNENGRGLSQDLNWCQJIAKSGYLEDVRLY 842
777 CRLIGICTSTVQLVQLMPYGLDHNENGRGLSQDLNWCQJIAKSGYLEDVRLY 836
843 HEDLAARNLVKSPNHKITDPEGLARLIDETEHYADGKPIKMMALSTLRRTHQ 902
837 HEDLAARNLVKTPQHVKITDPEGLARLIDETEHYADGKPIKMMALSTLRRTHQ 896
903 SDVWSYGVTVWELMTGAKPYDGIAPARIPDLLEKGERLPDPICITIDVYMIYKCMID 962
897 SDVWSYGVTVWELMTGAKPYDGIAPASDISISILEKESRPPICITIDVYMIYKCMID 956
963 SECRPFREIYSEFSMARDPQRFVYIQ-NEDLGPASPLDSFFYSLEDDDMGLVAAE 1021

```

```

DB 957 ADSRPFREIYSEFSMARDPQRFVYIQ-NEDLGPASPLDSFFYSLEDDDMGLVAAE 1016
QY 1022 EYLVPOGFGFCPPAPFAGAGVHHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEGA 1081
DB 1017 EYLVPOGFGFCPPAPFAGAGVHHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEGA 1042
QY 1082 GSDVFDLGLGAAKGLQSLPTHDPSFLQRYSEDPTVLPSET--DGYPALTCSPQPEY 1139
DB 1043 TSN-----NSTVACINRNGSCRYKEDATLQRYSSDPGTAVMEDNIDDAFL-----PVPEY 1092
QY 1140 VNQPDYVPPSPRREGGLPARPAGATLEAKTILSPKNGVYKDVFAFGAVENPEYL-T 1198
DB 1093 VNQ-SVPRRPAQSVQNVYHNQPLHP-----APGRDLHYQN--PHSNVAGNPEYLT 1141
QY 1199 PQGGAAPQHPHPAPFSPAFDNLVYWDQ-----DP-----PERGAPPTSTFKGPT 1242
DB 1142 AQ-----FTLSSGGRFNSPALMIKSGHMSLNDPVDQGFPPKETRNNGIFKG-PT 1191
QY 1243 AENPEYGLDVP 1254
DB 1192 AENAEYLRVAP 1203

RESULT 4
Q9YH40 PRELIMINARY; PRT; 1165 AA.
ID Q9YH40
AC 01-MAY-1999 (TREMURel. 10, Created)
DT 01-OCT-2000 (TREMURel. 15, Last sequence update)
DE 01-JUN-2002 (TREMURel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
GN XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RX MEDLINE=96241172; PubMed=9582016;
RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA Altschmid J., Schartl M.;
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT overexpression and mutational alterations.";
RL Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Schartl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53471; AAD10500.2; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DCM; 2.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.

```

SEQUENCE 1165 AA: 129614 MW: 777E38D8771A74E CRC64;

Query Match 39.3%; Score 2685; DB 13; Length 1165;
 Best Local Similarity 45.3%; Pred. No. 3.2e-196;
 Matches 579; Conservative 161; Mismatches 386; Indels 152; Gaps 32;

1 MELALCRWGLLALPPG-AAST-----OVCTGTMKRLPASPETHLMRLHLYQGQCV 55
 4 LELBEL-----LHLLILISIRCCSTDPDRKVCQGISNOMTM---LDHNYLKKMKMSGCNV 56
 56 VQGNLELYLPTNASLPLQDIOEVGVYVLAHNOVQVPLORLIRVSTQLFEDNYALA 115
 57 VLENLEITYEOENODLSFLOSIQEVGVYVLAHNEVSTPLVNLRLIRQNLYEENFTLL 116
 116 VLDNGPLNNTPTVYGASPGSLRELQSLTELKSGVLIQGNPOLCVDITLMDIRHK 175
 117 VMSNYOK-PPSSP--DVYVGLKQLSLNLTSLIGGVKVSHPNPLCNVETIMWMDIYDK 173
 176 NNQLALLTIDTRSRACHPCSPCKSGRCMGSESDCGLTRTVAGGC-ARCKGRLPTD 234
 174 TSNPTNMLPFAFERCCQCDPGCVNGSCWARGPHGCKFTLLCAEQCNRCGRPKPID 233
 235 CCHQCAAGCTGPKKSDCLACHFNHSGICELHCPALVTYNTDTESMPNPGRYTFGAS 294
 234 CCHQCAAGCTGPKKSDCLACHFNHSGICELHCPALVTYNTDTESMPNPGRYTFGAS 293
 295 CVTACPPNYLSTDVSCSLVCPLNQEVTAEDGTQRCESCKPCARCVGLM-----QYI 350
 294 CVKEGDSNVYTE-GACVRCGACMLEVD-ENGKASCRCPCDGVCKVCDGIGISLMTI 351
 351 KANSKFIGITTELFAGCKKIFGSLAFLESPFGDPSANTAPLQPELOLVFETLEITLYL 410
 352 AVNSTNIG---SFNCTKINGDILNRSFEGDHYKIGPMDEPHLNLTTVKIITGYL 407
 411 YISAMPDLPDLSVQNLQVITGRILHNGAYS-LILOGLISWLGIRSLREIGSLALIH 469
 408 VIMWPEMNTSLSVQNLLEIRGRITFSRGSFVAVVQVSHLOMGLSLKESASANVILK 467
 470 HNTLCFVHTVPMQDLFRNPHQALHTANRPEDECVGGLACHQACAGHCGPPTQCV 529
 468 NTPOLRYASTIMWRLLFRSEDSIEVDART-----ENQTNMNCSEDCGMBPPTNCV 520
 530 NCSQFLRGQECVECRVLQGLFREYVNAHRLCPHBECPQNGSVTCGPEADQCVACAH 589
 521 SCHAVDRGRCAVASCNLLQGEFREQVDRQVQCHQCELVQDLSITCYGPPRANSCKAH 580
 590 YKDPPECVABCPSPGVKPPDLSYMPWKPFDEBAGACQPCINCHSGVDLDDGCPAE--QR 647
 581 FODGPOCTPRCPHGMLSGDTL-IXKYADKMGQCCQPCONCTQGSGLSGCRGDIVSH 639
 648 ASPLTSIFNNFTVSFWLRVPKVASLHLEKRRQOKIRKXTMRLLQETELVEPLTPSGAMP 707
 640 SSLAVGLVSGLLITIVALLIV--LRRRRIR-KRRTIRRLQEKELVEPLTPSGAP 695
 708 NOAOWRILKETELERKYLKYGAGFTVYKGIIMPGENVAKIPVAIKULRENTSPRANKET 767
 696 NOAFRIILKETELERKYLKYGAGFTVYKGIIMPGENVAKIPVAIKULRENTSPRANKET 755
 768 LDEAYVAGVSPYVSRLLGICLSTVQVLTQMPYGLDLHVENRGRLSODLNMCM 827
 756 LDEAYVAVSDHPCRLGLGICLSTVQVLTQMPYGLDLHVENRGRLSODLNMCM 815
 828 CIAKMSYLEVRLVHRDLAARNVLYKSPNHYKTTDFGLARLLDIDETEHADGKVPK 887
 816 CIAKMSYLEVRLVHRDLAARNVLYKSPNHYKTTDFGLARLLDIDETEHADGKVPK 875
 888 MALESILRRFTSHQSDVWSYGVTVWELMTFGAKPYDGIIPARETIDLEKERLPOPPIC 947
 876 MALESILQWYTHQSDVWSYGVTVWELMTFGSKPYDIPAKETASVLENERLPOPPIC 935
 948 TIDVTMIVKCMIDSECRPRFRLVSEFSRMAPDQRFVVIQNEIDGLPASPLDSTFYS 1007
 936 TIEVMTILKCMIDPSSRPRFRLVSEFSRMAPDQRFVVIQNEIDGLPASPLDSTFYS 992

1008 LLEDDMGDLVDAEEYLVPOQGFPCDDPAPGAGVHRRSSSTRGGDLTLGLPSE 1067
 993 LLSDD--DVVDADEYLL-----RYKRIN-RQGS----- 1018
 1068 EEARSPPLASEGAGSDVFGDDLMGAAGLQSLPTDPSLQRYSEDPTV-PLPSETDG 1126
 1019 -----EPCIFPNHG-----PVRENSTALRYISDPTQNALKEKLDG 1053
 1127 YVAPLTCSPQPEYVNOQDVRPQ-----PSPRE-----GPLP-AARPAATLERAKTL 1173
 1054 H-----EYVNOGSESSRLSDIYNPNVEDLIDMGKPVLSQGEATNRSREYTL 1103
 1174 SPGNQGVKQVFAFGAVENPEYITPPCGAAPQCPHPAPSPAFDNLVWDQDPERGAP 1233
 1104 NTNQNSL---FLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGAL 1138
 1234 PSTFKGPTAENPEYLG 1251
 1139 TGNMGFLPAAENLEYLGL 1156

RESULT 5
 ID Q9W6F6 PRELIMINARY; PRT: 1137 AA.
 AC Q9W6F6
 DT 01-NOV-1999 (TREMURel. 12, Created)
 DT 01-NOV-1999 (TREMURel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMURel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN ERBB4.
 OS Gallus gallus (chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDBRNIN;
 RX MEDLINE=99263203; PubMed=10328894;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neurotrophin-1 (ng1) and erbB4 transcripts in
 RT embryonic chick hindbrain."
 RL Mol. Cell. Neurosci. 13:237-258(1999).
 DR EMBL: AF121963; A031764.1; --
 DR HSPB; P11362; IFCX.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001368; TNFR_c6.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; Kinase; 1.
 DR Pfam: PF01030; Recep_L domain; 1.
 DR Pfam: PF02757; YLP; 2.
 DR PRINTS: P00109; TYRKINASE.
 DR Prodom: PD000001; Euk_kinase; 1.
 DR SMART: SMO0261; FY3.
 DR SMART: SMO0219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 DR PROSITE: PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KM Kinase; Tyrosine-protein kinase.
 FT NON_TER

SEQUENCE 1137 AA: 127927 MW: 4D616436F87DC84F CRC64;

Query Match 38.7%; Score 2646.5; DB 13; Length 1137;
 Best Local Similarity 46.5%; Pred. No. 2.7e-193;
 Matches 532; Conservative 161; Mismatches 357; Indels 89; Gaps 26;

161 LCTQDITLMDKIDFKHKNQALLTIDTRSRACHPCSPCKSGRCMGSESDCGLTRTVG 220

```

Db      3 LCFADITHMDIYRNWASNTLVPTNGSSGCGRCHSKCTG-RCMPTENHCOITLKYVC 61
Qy      221 AGGC-ARCGPLPTDCCHEGCAAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDTF 279
Db      62 AECQDRCYGFYVSDCHRECAAGCGSPKDTDFACMNFNDSCACTGCPQFFVNPPTTF 121
Qy      280 ESMNPDEGRYTFGASCVTACPRVYLTSDVGSCTLVCPLEHNOEYTAEDTQRCCKSKPCA 339
Db      122 QLEHNNHAKTYTCAFVCKCPHNFV-VDSSSCVACPSKMEV-BENGIMCKPCTDTCF 179
Qy      340 RVCYGLGMOYIKANSKFIGITTELE-FAGCKIFGSUAFPESEFGDPASNTAPLOEBOLO 398
Db      180 KACDGIQTGSL-USAQTVDSNNIDKFINCTKINGNIIPLVTGHDHPYTHIAINPEKLN 238
Qy      339 VFETLEITGYLISAMPDLSLVSFQNLQVIRGLILNGVASLTLQGLGSMIGRL 458
Db      239 IFQVREITGYLISQWPEMNTDFRVFSNLVTIGGALYSGLSLILKQOGLTSLQFSL 298
Qy      459 RELSGGLALIHNTLCPVHTVWDOLFENPHQALHTANRPEDECVGEGLACHOLCARG 518
Db      299 KQISAGNIYITDMSNCYHTVWMTSLFSTPSQKVIHNNKKAENCTADGWCNCLCSD 358
Qy      519 HCGPPTQCVNSQFLRGECEVEGRVLOGLPREVNAHCLPCHPEQCP-QNGSVTF 577
Db      359 GCGPDPDCLSKREIRGRTICESCNLVDGEFREANSGVCECDPQCKEMDMWITCY 418
Qy      578 GPADQCVACAHKDPFPCVACRPSGVKPDLSYPIKPFDEBACQPCPINCSTHCVLD 637
Db      419 GCPDHCTCFHFKDQPCNCEKCPDLQGANSE-IFKXADEBRECHPHNCTQGC--- 473
Qy      638 DDGXCPEBQRA--SPLTSIFNNFTVSFWLRVPKVAS-----HLEKRRQ 680
Db      474 --RG-FASHDCIYFPT--ROSTLPQHATPPLIAGVIGLFIIVIMGLFAVVYRRKS 527
Qy      681 KIRKYMRLLOETELVEPLTPSGAMPNOAKMLKETELRKVYKLGSAFGVYKGM 740
Db      528 IKKRALRRL-ETELVEPLTPSGTAPNOQKILKELKRYKLGSAFGVYKGM 586
Qy      741 PDGENYKIPYALVLENTSPKANKELDEAVYVAGVSPYSLIGICTSTVOJLTOL 800
Db      587 PEGETKIPYALVLENTSPKANKELDEAVYVAGVSPYSLIGICTSTVOJLTOL 646
Qy      801 MPYGLLDHYRENRGLSGODILNMCQIAKMSYLEDVRLVHDLAANVLKSNHIX 860
Db      647 MPYGLLDHYRENRGLSGODILNMCQIAKMSYLEDVRLVHDLAANVLKSNHIX 706
Qy      861 ITDFGLARLLDIDETEHADGKVPKIMWALSLRRRTTHOSDVMSYGVTWELMTFGA 920
Db      707 ITDFGLARLLDIDETEHADGKVPKIMWALSLRRRTTHOSDVMSYGVTWELMTFGA 766
Qy      921 KPYDGIPTAREIPULLEKGERLPQPICTIDVYIMYKCMKIDSECPREELVSEFSMA 980
Db      767 KPYDGIPTAREIPULLEKGERLPQPICTIDVYIMYKCMKIDSECPREELVSEFSMA 826
Qy      981 RDQRFVYIIONED-LBPASPLDSTFRSLLEDMDMDLDAEYLVPOQGFPCPDAPGA 1039
Db      827 RDQRFVYIIONED-LBPASPLDSTFRSLLEDMDMDLDAEYLVPOQGFPCPDAPGA 885
Qy      1040 GGVVHRRHSSSTRSGGDTLTGLFSEEBAPRS--PLAP-SEGASDVFDGDLGMAAK 1096
Db      886 RTRIDSNRNQFVYRDGYAAEQGV-FMYPAPFCIIPEAPVAGAAFAEFEDTCCNGTLR 944
Qy      1097 GLGSLPTDPSPLQRYSEDPVPLPS-----ETDGYVAPLTCSPQRYVNOQPDVRRP 1149
Db      945 KQVATLAKEDSSQRIQSADPTVFIPIRVIRGELDEDEGYWTPEMDKXTDYLMPVEENPV 1004
Qy      1150 PSREGLPLPA-RPAGATLERAKTISPKNQVYKVF-----AFGAVENPEYLTPO 1200
Db      1005 SRKNGDLQAVNDPEYHN-----APNGQPKADEYVNEPLVNTFAUTLEAEYL--- 1054
Qy      1201 GGAAPQHPPPAPSPAFDNIYVWDQDPPERGA--PSTFKGTPT-----AE 1244

```

```

Db      1055 -----KNULPEKAKKAFQNDPDYKNSLPPRETLQHPDYLGEYTFYKQNGRIRPIVAE 1109
Qy      1245 NPEYL 1249
Db      1110 NPEYL 1114

RESULT 6
P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TRENBLREL. 03, Created)
DT 01-MAY-1997 (TRENBLREL. 03, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Erib3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gelliner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes."
RL Genome Res. 9:251-258(1999).
EMBL AF056116; AAC34391.1;
HSSP; P11362; IFCX.
InterPro; IPR000494; EGFR_L_domain.
InterPro; IPR000719; Euk_kinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_kinase.
Pfam; PF000257; Furin-like; 1.
Pfam; PF00069; Pkinase; 1.
Pfam; PF01030; Recep_L_domain; 2.
ProDom; PD000001; Euk_kinase; 1.
SMART; SM00261; FU; 3.
DR SMART; SM00219; TYRKC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647B9 CRC64;

Query Match 33.1%; Score 2260; DB 13; Length 1328;
Best Local Similarity 39.9%; Pred. No. 1.2e-163;
Matches 516; Conservative 155; Mismatches 418; Indels 204; Gaps 33;

9 WGLLALLP--GAASD---VCTGDMKRLPASPEPTHLDMLRLHYQGVYQGNLEL 62
4 WRLTIMCVASRLRAASGOTQAVCPGTONGLSSTGSENOYNLMDRYKGEIIMGNLEI 63
63 TYLPTNASTSLFLODIOEYQGVYLAHNOVRQVPLQRIYVGTOLFEDNYALAVLNDGP 122
64 TOIESNMFSLTKIRRETVGLVLANHFGQILPQGLRVIRGNSLYERRFLSFLN--- 120
123 LNTTPTVTGASPGGLRELQSLRLTEILKGVYIQRPOLCYQDTILMKDIFPKXNQLALT 182
121 ----YPKDG--PSGLNQGLMNLTEILLDGVQIINNYLRXGPMVYRRDII-RNNDPIE 173
183 LIDTMSAACHPCSPMCGSRGWSSESDQSLRRTVCAGGC-ARCGPLPTDCHBOCA 241
174 IQNGERGVCH--KSC-GNVCWPGKDOQILTKIVCAPOQNDRCFTGSPRDCHECA 229
242 AGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDTEESMPNDEGRYTFGASCVTACPY 301
230 AGCGPLDPTDCAFRLFNDSGACVPCQGLIYKQFFQWETNNAYQVSGISCVSCPT 289
302 NYLSTDVSSCTLVCPLEHNOEYTAEDTQRCCKSKARCYGLGMOYIKANSKFIGIT 360
290 HFV-VDGSSCVSVCPDKMEV--ERGSQROCELCGLCPKVCETGAE---QRTVDSS 342

```

```

QY 361 ELE-FAGCKKIFGSLAFLPESFDDGASNTAPLQPEQLQVFFETLEITGVLTVISAMPDSL 419
DB 343 NIDSFNCTKIQSLHFLVTGLIGDPPKVPPLDADKLEFPRVITITILITQSWPKL 402
QY 420 PDLSEFQNLQVIRGILHNGAVSLTLQGLISWLGRLSRELSGSLALIHNTLHLCFVHT 479
DB 403 NDLVSFSLTLTGRLSLFKRFLMWRIPITLTLGLRLSRLSDGSVYISQNAHLCHHT 462
QY 480 VPMDFLFRNH-QALLHTANRPEDECVGSLACHQICAGHCHGKPGPPTCCVNCQFLRQ 538
DB 463 VMTQLFRGSRVANSINRPMACVAGRCVDCPSGSGGPGPDCLSCRNYSRHG 522
QY 539 ECVEECRVLQGLFREYVNAARH-CLPCHPEQOPONGSVTCFPGPADCCVACAHYKDPFV 597
DB 523 TCAGCHFNHSGIRERFAGLNGVVAACHPECKPOTGASCTGPADECMCTFRDGPYCM 582
QY 598 ARCPGVKPDLSYMPIWTFPEDEGACPPCINCTHSVDLDDKGPACGRAS----- 649
DB 583 SSCPAGVN-DGEKGLIFKFPNREGHCPCQONCTQCSGPGINDCLARLTSSQITG 641
QY 650 -----PLTSIFNNFTVSFMLRVKVASHLKRRQOKIRKYMRLLOTELVEPLTBSG 704
DB 642 IALGVPAGLIF-CLVLFPLGM-----LYHGLAIRKRRANRYLESSESPEPLGP-G 691
QY 705 AMNQOAMEILKETEELKVKVLSGSAFGTVYKGIWIPDSENYKIPVAKYLRNENTSPKAN 764
DB 692 EKGTKVHAILKRSDLRKIKPLGSGVFTVSGFWMPEDEYKIPALIKTIDSSGRQIF 751
QY 765 KELDEAYVMAGVSGSVYVRLIGICTSTVOLVTOCMYPCGLLDHYRENRGLSGDILIN 824
DB 752 TETTHLLMSGSLDHYIVRLDGLICGTLOLQVTOSSGSLLEHRRQHTSLDQRLIN 811
QY 825 MCMQIAKMSYLEDEVLVHRLDAAARNVLKSPHNVKITPFGARLIDIDETVYHDDGKV 884
DB 812 KCVQIAKMYVLEENHVAKNLAAARNILKANDYQOISYVADLLYPDKRYVSEYTKT 871
QY 885 PIKMALESILRRFTHQSDVMSYGVTTWELMTFGAKPYDGI-PAREIPDLLEKGRLP 944
DB 872 PIKMALESILRRFTHQSDVMSYGVTTWELMTFGAKPYDGI-PAREIPDLLEKGRLP 931
QY 945 PCTIDVYMWKCMWIDSECRPRPREIVSEFSRMAQRQVVIQNEGLGASVLDSTF 1004
DB 932 ALCTIDVYMWKCMWIDSECRPRPREIVSEFSRMAQRQVVIQNEGLGASVLDSTF 980
QY 1005 YRSLLEDDMDGLVAEEYLVPOGFFCPDPAFGAGVHNRHSSSTRSGGDLTLC 1064
DB 981 -----EDSGMGEL-----RGSER-----GLEADLE 1003
QY 1065 PSEERARSPPLAPSEGASDVFDGLGMC---AAGGLQSLPTHDSPLQ-----R 1111
DB 1004 EDEEB-----GLGDRFATPSLOPSPSWSTSPQINSYVMTOLR 1042
QY 1112 YSEDPVPLPSETDGVAPLTCSPOP-EYVNO-----PVRPQPPSPRGP 1156
DB 1043 VD-----FVSGGHIGIYLPMSFSPVDITRQLMQORSLSSVRLPDRSAFRSSRAE 1096
QY 1157 L-PAARPAAGATLEPAKTLSPGKNGVVDVAFGAIVENPEYLTQGGAAAPPHPPAFS 1214
DB 1097 LCEDGAQCAGIFRVR-----FGSERGN-----FOGG----- 1122
QY 1215 PAFDNLVYWDODPPERGAAPTFTKGTPTAENPE 1247
DB 1123 -----QQRKLTASSPSSFKTWADEDE 1146

```

RESULT 7

Q9BIH9 PRELIMINARY: PRT: 1433 AA.

AC Q9BIH9: 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Putative epidermal growth factor receptor (Fragment).

```

GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_Taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SLA.
RA Lycet G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ01655; CAC35008.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recp_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00113; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM Receptor.
FT NON TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.1%; Score 1988.5; DB 5; Length 1433;
Best Local Similarity 32.4%; Pred. No. 8.2e-143;
Matches 469; Conservative 196; Mismatches 397; Indels 385; Gaps 38;

```



```

QY 489 PHQALHTANREDECEVGEGLACHOLACAGCHWGPPTQCVNCSQFLRQGECEVGRVLD 548
DB 450 SDHEVWVQKRNATTECHHEEMECSEGCSSAGCKGKPEQCEKRVKXKGLDCK--- 506
QY 549 GLPREY-VNARHCLPHGPCQFONGSVTCGPEADCCVACAHYKPPFCVACP----- 601
DB 507 SLFRLYSVDSKTGDCHQOECKD-----FCYGNEDCGSCMVNKKDRFCVACCPPTTKAM 561
QY 602 -----SGVPDLSYMPIMKFPD----- 618
DB 562 NGCTCNCHKTVCGRGPRDTIAPDGCISCDAKIGSDAKIERGLMDESCPDGYSIDVL 621
QY 619 -EEG----- 621
DB 622 QEEGPLKQSGKAVCRKCRCKCTGYGFHEQFOECTGYKKGECECECFDFAANE 681
QY 622 -ACQCPINCT-----HSCVDL-----DD-----KCCPME-----QPA 648
DB 682 TRICLPCHQECRGCHGLGDDHHCENLKLFEQDPYDNTFTTCVSNCPKSHPKRPPQEA 741
QY 649 SPLTSIFNFTVSPFWLRV-PKVS-----ASHLEKRRQOKIRKXTM 687
DB 742 GKICPYCASDMSGSLRIEPOQTQVITWGSVMALLLCVFGIAFLFRHKKKDAVVM 801
QY 688 RLLQETELVEPLTPSGAMPNQAOMRLKTELKXKXVUGSAGFGVYTKGIMPGEANV 747
DB 802 TMLACCESEPLRPANVGPNLTKRIKEAEIRRGVGMGAFGVFGVWMPGESEVX 861
QY 748 IPVAIKVIRENTSPKANKEILDEAYVAVAGSVYRLLIGICTSTVQVLTQMPGCLL 807
DB 862 IPVAIKVIRENTSPKANKEILDEAYVAVAGSVYRLLIGICTSTVQVLTQMPGCLL 921
QY 808 DHVRENKRGISGOLLNMCQIAKGSYLEDVFLVHDLAARNLVKSPNHYKITDFGLA 867
DB 922 DVARNKXKXIGSALLNWSQIARMAVLEERLVRDLAARNLVQTPSCVKITVFGIA 981
QY 868 RLIDIDETEYHAGGVPIKMALESILRRFTHSDVSYGVTVELMTFGKPPDGP 927
DB 982 KLDFDSDEYRAAGKMPIMLALBCTRHRVFTSKSDVMAFGITIMELLITYGARPYENP 1041
QY 928 AREIPDLLEKGERLPOPPICITIDVIMVCMIMIDECRPRELVSERSMARDPQREV 987
DB 1042 AKDVPELIEIGHKLPOPDICSLDLYCILSCVLDADARTFKQLAETFAKARDPGRYL 1101
QY 988 VIONEDUGPASPLDSTFYRSLLEDDMDGLV----- 1018
DB 1102 MI-----PQDKFWRLPSYTNQDEKDLITLAVMAAAAAAGASNVDPSTIA 1152
QY 1019 DAEYIVPQCGFCPPAPAGAGVHHRSSSTRSGGDLTLGLPSEEAARS----- 1073
DB 1153 EIDEXILOPKTRPSIMLPESA-----VEPS-DEMPKSLRYCK 1188
QY 1074 -PLAP---SEGASDVFDGDLGMAKGLQSLPTHDPSPLOKYSSEPTVPLPSETDGYVA 1129
DB 1189 DPLKPDDETIDHGEV-----GVGJIR-----INTLPLDEDDYLM 1222
QY 1130 PLTCSPOPEYVNOQVDRPQPPSPREGFLPARPAGATLERAKTILSPGXGVVDPVAFGG 1189
DB 1223 P-TCQSO-----NOS-----TBQ-----YMDLGVTA 1243
QY 1190 AVENPEYL-----TPGGAAPOPHPPPAPAFDNLVYWDQDPPERGAPPTFKGT 1240
DB 1244 SVNPEYLMGSTQAIAGLAGSWG--PHTPP-----PPTNPGM 1280
QY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

RESULT 8
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
Q9UK79;

```

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RL autoinhibitor."
RN [2]
RP Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;
RT Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
RL EMBL; AF177761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FCI1BR347E2D030C CRC64;

Query Match 27.4%; Score 1871; DB 4; Length 419;
Best Local Similarity 96.8%; Pred. No. 1.2e-114;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELALCRWGLLALLPPGAASVQCTGTDMKRLPASPEHLDMLRHLVGGCQVQGNL 60
DB 1 MELALCRWGLLALLPPGAASVQCTGTDMKRLPASPEHLDMLRHLVGGCQVQGNL 60
QY 61 EUTYPTNASHSLFDIOIOVGVYLIANOVQVPLQRLRVRGTLFEDNYALAVLDNG 120
DB 61 EUTYPTNASHSLFDIOIOVGVYLIANOVQVPLQRLRVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTAVTGASPGGLRELQRLSTELILKGVYIQRNPOLCYODTILMKDIFHNQOLA 180
DB 121 DPLNNTTAVTGASPGGLRELQRLSTELILKGVYIQRNPOLCYODTILMKDIFHNQOLA 180
QY 181 LTLIDTNRSRACHPCSPKCKSRCKWGESSEDCQSLTRTVCGAGCARCKGPPPTCCHQC 240
DB 181 LTLIDTNRSRACHPCSPKCKSRCKWGESSEDCQSLTRTVCGAGCARCKGPPPTCCHQC 240
QY 241 AAGCTGPHNSOCLACLNHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
DB 241 AAGCTGPHNSOCLACLNHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
QY 301 YNYISTDVGSCTLVCFILNQEVTAEADGTQRCCKSKPCARVYCYG 345
DB 301 YNYISTDVGSCTLVCFILNQEVTAEADGTQRCCKSKPCARVYCYG 345

```

RESULT 9
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
SQ Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.


```

Db      141 PEETATPKTGP--DHCKMCAHFIDGHCVKACPAVGLGENDTL-VWKYADANAACOLCHP 197
Qy      629 NCTHSCVDLDDKGCAPASPLTISFNNFTVSFWLRYVKVASHLEKROOKIRKYTMR 688
Db      198 NCTRGCKGPGLECCP---NGSKTPSIAGVVGGLCLVVGIGIGLYLRRIIVYKRLR 254
Qy      689 RLLOETELVEPLTPSGAMPNOAMRLKETELRKVKVLSGAFGVYKGIWPDGENVKI 748
Db      255 RLLOERELVEPLTPSGEAPNOAHLRIKETEFKVKVLSGAFGVYKGIWPDGENVKI 314
Qy      749 PVAIKYLRRENTSPKANKELIDEAYVWAGVSPYVRLIGICTSTVQVLTOLMPYGCILD 808
Db      315 PVAIKELRATSPKANKELIDEAYVWASVDNPRVCLGICTSTVQVLTOLMPYGCILD 374
Qy      809 HVRENKRGRLGSDLLMCMQIAKMSYLEDVLRVLRDLAARNVLYKSPNHVKITDFGLAR 868
Db      375 YIREKMDNGSQVLLMWCQIAKGMVYLERLVRDLAARNVLYKTOHVKITDFGLAK 434
Qy      869 LLDIDETEHADGGKVPFKMALESLIRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 928
Db      435 LIGADEKEVHAEGKVPFKMALESLIRIYTHQSDVWSYGVTVWELMTFGSKPYDGI 494
Qy      929 REIPDLKEGERLPOPICTIDVYIMVCMMDSECRPRFRELVSFSPMARDPQRFVY 988
Db      495 SEISSYLEKGERLPOPICTIDVYIMVCMMDADSRKRFELIAESKARDPRYLYV 554
Qy      989 IQ-NEDLGPASPLDSTFFYSLLEDMDGDLVDAEYLYPQGFPCPDPAAGAGVHHRH 1047
Db      555 IQDERMHLPSPDTSKFRYRLTMEEDMEDIVDAEYLYPHQGF----- 598
Qy      1048 RSSSTRSGGGDLTLGLEPSEEEAPRSP-----APSEGGSPVFDGDLGMAKGLSLP 1102
Db      599 NSPST-----SRPPLSLSLATSNSNATNCID-----KNGQCHP 632
Qy      1103 THDPSPLOGYSDPTVPLPSET--DGVAAPLTCSPQPEYVNOQDVPAPSPREGPLPA 1160
Db      633 VREDSFVQYSSDPTGPNFLEESIDGFL-----PAPEYVNO--LMPKXPS----- 675
Qy      1161 RAGATLERAKTLSPCKNGVVDV-----AFGAVENPEYL 1197
Db      676 -----TAMVQNIYNNISLTAISKLPMSRYONSHSTAVDNPEYL 715

RESULT 12
086714 PRELIMINARY: PRT; 567 AA.
ID 086714;
AC 086714;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCB1_Taxid=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoshcheck C., Janeson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities."
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase.1.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase.1.
DR SMART; SMO0219; TYRC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

```

```

DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON TER
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD725E1 CRC64;

Query Match
Best local similarity 54.6%; Score 1673; DB 15; Length 567;
Matches 351; Conservative 66; Mismatches 126; Indels 100; Gaps 13;

Qy 578 GPEADQVCAAHYKDPFCYACRCBSGKPLSLMTFKPFDEGACQPCPCINCHSCVDL 637
Db 1 GP--DHCKMCAHFIDGHCVKACPAVGLGENDTL-VWKYADANAACOLCHPNCRGCKGP 57
Qy 638 DDKGCAPASPLTISFNNFTVSFWLRYVKVASHLEKROOKIRKYTRRLLOETELV 697
Db 58 GLECCP---NGSKTPSIAGVVGGLCLVVGIGIGLYLRRIIVYKRLRLLOERELV 114
Qy 698 EPLTPSGAMPNOAMRLKETELRKVKVLSGAFGVYKGIWPDGENVKI PVAIKYLR 757
Db 115 EPLTPSGEAPNOAHLRIKETEFKVKVLSGAFGVYKGIWPDGENVKI PVAIKELRE 174
Qy 758 NTSFKAKELIDEAYVWAGVSPYVRLIGICTSTVQVLTOLMPYGCILDHVENRGR 817
Db 175 ATSPKAKELIDEAYVWASVDNPRVCLGICTSTVQVLTOLMPYGCILDYIREKMDNI 234
Qy 818 GSQDLNMCQIAKMSYLEDVLRVLRDLAARNVLYKSPNHVKITDFGLARLLDIDET 877
Db 235 GSQYLLMWCQIAKGMVYLERLVRDLAARNVLYKTOHVKITDFGLAKLIGADEKE 294
Qy 878 HAQGGKVPFKMALESLIRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREPDL 937
Db 295 HAQGGKVPFKMALESLIRIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASEISYLE 354
Qy 938 GERLPPOPICTIDVYIMVCMMDSECRPRFRELVSFSPMARDPQRFVYIQ-NEDLGP 996
Db 355 GERLPPOPICTIDVYIMVCMMDADSRKRFELIAESKARDPRYLYV IQDERMH 414
Qy 997 ASPLDSTFFYSLLEDMDGDLVDAEYLYPQGFPCPDPAAGAGVHHRHSSSTRSGG 1056
Db 415 PSPDTSKFRYRLTMEEDMEDIVDAEYLYPHQGF-----NSPST----- 454
Qy 1057 GDULTLGLPESEEEAPRSP-----APSEGGSPVFDGDLGMAKGLSLPTDPSPLO 1111
Db 455 -----SRPPLSLSLATSNSNATNCID-----KNGQCHPVREDSFVQR 492
Qy 1112 YSDPTVPLPSET--DGVAAPLTCSPQPEYVNOQDVPAPSPREGPLPAPRAGATLER 1169
Db 493 YSDPTGPNFLEESIDGFL-----PAPEYVNO--LMPKXPS----- 526
Qy 1170 AKTISPCKNGVVDV-----AFGAVENPEYL 1197
Db 527 -----TAMVQNIYNNISLTAISKLPMSRYONSHSTAVDNPEYL 566

RESULT 13
064895 PRELIMINARY: PRT; 962 AA.
ID 064895;
AC 064895;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN Gag, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCB1_Taxid=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
RA "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B";

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL; X52209; CA36459.1; -
 DR EMBL; X52211; CA36459.1; JOINED.
 DR HSSP; P10828; ZNLL.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR000535; Hormone_rec_119.
 DR InterPro; IPR001723; Stcdhmm_receptor.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR001628; Znf_C4steroid.
 DR Pfam; PF00104; hormone_rec; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00105; zf-C4; 1.
 DR PRINTS; PR00398; STRDHOMONER.
 DR PRINTS; PR00047; STROIDFINGER.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR ProDom; PD000035; Znf_C4steroid; 1.
 DR SMART; SM00430; HOL1; 1.
 DR SMART; SM00219; TYRKC; 1.
 DR SMART; SM00399; Znf_C4; 1.
 DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KM Transcription regulation; Tyrosine-protein kinase;
 KM Zinc-finger.
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B5C6 CRC64;

Query Match 23.5%; Score 1608.5; DB 15; Length 962;
 Best Local Similarity 50.9%; Pred. No. 5.2e-114;
 Matches 352; Conservative 66; Mismatches 156; Indels 117; Gaps 16;

QY 541 VEECRVLQGPREF-YNNAR-HCLP-----CHPEQC 568
 DB 354 IEKQSESYLAEHYNYRGNIPHWKILMKVADLMIGAYHARFLMKVECPTELS 413
 QY 569 PONGSYTCGPEADQCVACHYDPPFCVACPSGVKPDLSYMPKPFDEGACQPCPI 628
 DB 414 PQE-----VGP--DHCKMCAHFIDGPHCVKACPAAGVGENDTL-VKRYADANAVCOLCHP 465
 QY 629 NCHSCVDDDKCKCPAEPQASPLTSFNNFTVSFWLRVPKVSASHLEKROQKIRKYTKR 688
 DB 466 NCHRGCKGPGLECCP--NGSKTPSTIAGVVGGLCLVAVGGLGIGLYRRRIHVRKRTL 522
 QY 689 RLQETELVEPLTPSGAMPNOAQMRILKETELRKVKVLSGAGFTYKXIMIPDGENVKI 748
 DB 523 RLQGERELVEPLTPSGEAPRQAHRLIKETEFKVKVLPFGAGVYKGLMIPBEGKYTI 582
 QY 749 PVAIKVLRNTSFKANKELIDEAYVAVGSPYVSRLLGICLTSTVQLVTQLMPYGCILLD 808
 DB 583 PVAIKELREATSPKANKELIDEAYVAVSDNPHVCRLLGICLTSTVQLVTQLMPYGCILLD 642
 QY 809 HYRENKRGLSODLLMCMQIAKMGSYLEDVRLYHNDLAARVLYKSPNHVKTITPGLAR 868
 DB 643 YIREHNDNGSQYLLMWCQIAKMGNYLEERHVNHDLAARVLYKTPDHVKTITPGLAK 702
 QY 869 LLDIDETEHADGKGVPIKMMALLESILRRFTHOSVMSYGVTVWELMFGAKPYDGIIPA 928
 DB 703 QCGADEKEYHAEGKGVPIKMMALLESILHRIYTHOSVMSYGVTVWELMFGAKPYDGIIPA 762
 QY 929 RLIPDLLEKGERLPQPICTIDVYIMVCMIDSECRPFRELVESEGRMARDPQRFVY 988
 DB 763 SEISSVLEKGERLPQPICTIDVYIMVCMMSGDSRPFRELLIAEFGRMARDPQRFVY 822
 QY 989 IO-NEDIGPASELSTFFYSLLLEDDMDGLVDAEEXVYVQGFCEPDPAFGAGVNHHRH 1047
 DB 823 IGDGERMHLPSPTDSGFYRTIMEEDMEIDVDAEYLPHVQGF----- 866
 QY 1048 RSGSTRSGGGDLTLGLSEPEEAPRSPAPSEGAGSDVFDGLGMAKAGLQSLPTHDP 1107
 DB 867 NSPST-----SRTPLSLSLSTSN-----NSATKCIDRNGCH--- 898

QY 1108 PLQRYSEDPYLPSEITDGYVAPLTCSPQPEYVNPQDVRFQPPSPRESGPLPAARPAAT- 1166
 DB 899 -----PVREDFGL-----PAPEYVNG--LMPKPSSTAMVQNGYVNSILSLA 937
 QY 1167 LERATLSPGKGVKVDVPAFGAVNPEYL 1197
 DB 938 ISKLPMDSRYQN-----SHTAVDNPEYL 961

RESULT 14
 Q85468
 ID Q85468 PRELIMINARY; PRT; 545 AA.
 AC Q85468;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Avian erythroblastosis virus (T833) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
 NX NCB1_Taxid=11861;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88217326; PubMed=2897102;
 RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation."
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL; X06943; CA30024.1; -
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TYRKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KM ATP-binding; Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DC8C8CA0FB4F4 CRC64;

Query Match 23.4%; Score 1600; DB 15; Length 545;
 Best Local Similarity 54.1%; Pred. No. 9.9e-114;
 Matches 339; Conservative 63; Mismatches 135; Indels 90; Gaps 13;

QY 578 GPEADQCVACHYDPPFCVACPSGVKPDLSYMPKPFDEGACQPCPICTHSCVDL 637
 DB 1 GP--DHCKMCAHFIDGPHCVKACPAAGVGENDTL-VKRYADANAVCOLCHPNCRGCKGP 57
 QY 638 DQKGPAPQASPLTSIFNNFTVSFWLRVPKVSASHLEKROQKIRKYTMRELLOETELV 697
 DB 58 GLECCP--NGSKTPSTIAGVVGGLCLVAVGGLGIGLYRRRIHVRKTLRLQERELV 114
 QY 698 EPLTPSGAMPNOAQMRILKETELRKVKVLSGAGFTYKXIMIPDGENVKIPVAIKVLE 757
 DB 115 EPLTPSGEAPRQAHRLIKETEFKVKVLPFGAGVYKGLMIPBEGEVITPVAIKELRE 174
 QY 758 NTSRKANKELIDEAYVAVGSPYVSRLLGICLTSTVQLVTQLMPYGCILLDHYENRRL 817
 DB 175 ATSPRANKELIDEAYVAVSDNPHVCRLLGICLTSTVQLVTQLMPYGCILLDYIREHNDI 234
 QY 818 GSQDLLMCMQIAKMGSYLEDVRLYHNDLAARVLYKSPNHVKTITDGLARLLIDREY 877
 DB 235 GSQYLLMWCQIAKMGNYLEERHVNHDLAARVLYKTPDVKTITDGLAKQGADEKEY 294
 QY 878 HADGKVPKMMALLESILRRFTHOSVMSYGVTVWELMFGAKPYDGIIPARBIPLLEK 937
 DB 295 HAEQKVPKMMALLESILHRIYTHOSVMSYGVTVWELMFGSKPYDGIIPASEISSVLEK 354
 QY 938 GERLPQPICTIDVYIMVCMIDSECRPFRELVESEGRMARDPQRFVYIO-NEDLGP 996
 DB 355 GERLPQPICTIDVYIMVCMMSDADSRPFRELLIAEFGRMARDPQRFVYIIGDERMHL 414

QY 997 ASPLDSTFYRLSDDDMDLVDAEYLVPQGFPCPPAPAGAGVYHRRSSSTRSG 1056
 DB 415 PSPTDSKFRYTLWEEDMEDIVADELYVHQGF-----NSPT----- 454
 QY 1057 GDLTLGLEPSEEPAPSP-----APSEAGSDVFPDGLMGAKGLQSLPHYDPSLOR 1111
 DB 455 -----SRTPILSSLSATSNNSATNCIRNG-----H----- 481
 QY 1112 YSEDPYVLPSETDGYAFLTCSPOPEYVNPQVDPQPSREGPLPAAPAGAT-LEBA 1170
 DB 482 -----PVEDGFL-----PAREYVNO--LMKKSTANVQIYINISLTASTL 524
 QY 1171 KTLSPKRGVVKDVFAGAVENPEYL 1197
 DB 525 PMDSRYON-----SHSTAVDNPEYL 544

RESULT 15

Q9WVF5 PRELIMINARY; PRT; 655 AA.
 ID Q9WVF5
 AC Q9WVF5
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 12, Last annotation update)
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schnell C.,
 RA Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Maibhe N.J.;
 RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-terminal truncated Receptors."
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH/101, 129/SVJ, AND 129/SVEVTC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampand A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maibhe N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Aikawa T., Hata A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Balderrelli R., Barsch G.,
 RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita K., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima Y., Mazzarelli J., Mombarts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Willing L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL, AF124513; AAD44149.1; -
 DR EMBL, AF275365; AAG28047.1; -
 DR EMBL, AF275365; AAG28047.1; JOINED.
 DR EMBL, AF275365; AAG28047.1; JOINED.
 DR EMBL, AK004944; BAB23688.1; -
 DR EMBL, AK004944; BAB23688.1; -
 DR EMBL, AK004911; BAB23662.1; -
 DR MGD, MGI:95294; Egfr.
 DR InterPro, IPR00494; EGFR_L domain.
 DR InterPro, IPR002174; Furin-like.
 DR Pfam, PF00757; Furin-like.
 DR Pfam, PF01030; Recep_L_domain; 2.
 DR SMART, SMO0261; FU; 3.
 KW Receptor.
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.0%; Score 1506.5; DB 11; Length 655;
 Best Local Similarity 44.5%; Pred. No. 1.8e-106;
 Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 LTLALPPGAA--STQVCTGDMKRLPASFTTHDMLRHYOGGVQVGLTYLPTN 68
 DB 14 LTLALCAAGALBEKKVCGTNSNRTQJGTFEDHLSLQRYNNCEVVLGNLITYVORN 73
 QY 69 ASLSFLQDIOEVQGYVLLIAHNVQVPLQRLRIYRGTLQFEDNYALAVLNDGPDPLNNTTP 128
 DB 74 YDLSLTKTIQEVAGVLLALNTVERIPLENQIIRGNALYNTALATLS----- 124
 QY 129 VTGASPGGLREQLRLSTELLKGVLLIQNPQLCYDTILMKDI---FHQNOALATLI 184
 DB 125 -YGTNRTGLRELPMNMLEILIGAVFSNPNILCNMDIQRVDIYQVFNMSMDL--- 180
 QY 185 DTRNRACHPSPMKGSGRCESESDQSLRTVYAGGCA-RCKGPLPTOCHEQCAAG 243
 DB 181 -QSHSSCPKCDSPNSCWGGEBENQKTKIKIQAQCHCRGRSPDCHQCAAG 239
 QY 244 CTGPHSDCLALAHFNHSGELCHCPALVTYNTDTFESMPNPSRYTFGASCVTACPNY 303
 DB 240 CTGPRSPDCLVQKFDATCKDTCPPMLNYPPTYQMDVNPCKYSGATCVKCKCPNY 299
 QY 304 LSTVGSCTVCPRLNCEVTADGTQREKSKCARCYGLQMYIANKSFGITILE 363
 DB 300 VTDHGSCTVACGPDYEV-BEDGIRKCKCDGPKKVCNIGIGEFK-DTLSINATNIK 357
 QY 364 -FAGCKKTFGLAFIPESFGDPPASNTAPLQPEQLQVETLEETGYLYISAMPDLPDL 422
 DB 358 HFKYCTAISGLHLIPVAFKXDSFTRTPPLDPRLEIKYKEITGFLIQAMPDNMTDL 417
 QY 423 SVFQNLQYIRGILHNGVSLTLQGLISWLGSLRLSGGLALIHNTHLCFVHTVPW 482
 DB 418 HAFENLEIRRTKHGQFSLAVVGLNTISGLSLKXISGDVYIISGNRLCYANTINW 477
 QY 483 DOLFENPQALHTANREDECVSGELACHOLCARGHWGGPFCVNCVCSFGLGOEVE 542
 DB 478 KKLFTPTQKTKIMNRKCKCAVNHVNCPLCSSEGWGPEDCVSCVNSGRREVE 537
 QY 538 KCNILEGPRFEVENSECICQHPCLPANNITCTGSPRDNICQAHYIDGPHCVKTCFA 597
 DB 543 ECRVLQGLPREYVNAHGLCPHPECOPNGSVTCFEPADQVCAAHYKDPFVCARCS 602
 QY 603 GVKPDLSTYMPKPPDEGACQPCPINTHSCVLDKGC 642
 DB 598 GIMGENNTL-VKRTADANNVCHLANCTYCGAGPGLQGC 636

Search completed: July 22, 2003, 09:24:19
 Job time : 51.8246 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	6627	97.0	1255	21	AA932620	Human heterogulin 2
2	6627	97.0	1255	22	AAE12130	Human tyrosine kinase
3	6627	97.0	1255	22	AAE60157	HER2 transgene pl
4	6627	97.0	1255	23	AAU74545	Human HER2 (ErbB2)
5	6623	96.9	1255	17	AAW01111	HER-2/neu protein
6	6623	96.9	1255	20	AAW92406	Human HER-2/neu p
7	6623	96.9	1255	21	AAAB21198	Human HER-2/neu p
8	6623	96.9	1255	21	AAV84780	Amino acid sequenc
9	6623	96.9	1255	22	AAAB8458	Human HER-2/neu p
10	6623	96.9	1255	22	AAAG8367	HER2/neu amino aci

[illegible]

QY 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780
 DB 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780
 QY 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEYDR 840
 DB 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEYDR 840
 QY 841 LVHRDLAARNVLYVSPNHVKTIDFGALRLDIDETEHAAOGKVPKIMMALESILRRFT 900
 DB 841 LVHRDLAARNVLYVSPNHVKTIDFGALRLDIDETEHAAOGKVPKIMMALESILRRFT 900
 QY 901 HQSDVMSYGVTTWELMTFGAKPYDGI PAREI PDLEKGERLPORPCTTIDVYIMYKCMW 960
 DB 901 HQSDVMSYGVTTWELMTFGAKPYDGI PAREI PDLEKGERLPORPCTTIDVYIMYKCMW 960
 QY 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGPASPLDSTFYRSLDEDDMDGLVDA 1020
 DB 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGPASPLDSTFYRSLDEDDMDGLVDA 1020
 QY 1021 EBYLVPOQGFECPPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEEA PRSPLAPSEG 1080
 DB 1021 EBYLVPOQGFECPPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEEA PRSPLAPSEG 1080
 QY 1081 AGSDVFPDGLCMGAKLOSLPTHDSPLOQYSDPTVP PSETDGVARLTGSPQPEYV 1140
 DB 1081 AGSDVFPDGLCMGAKLOSLPTHDSPLOQYSDPTVP PSETDGVARLTGSPQPEYV 1140
 QY 1141 NOPVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGAVENPEYLTPO 1200
 DB 1141 NOPVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQPPERGA PPTSTKGTPTAENPEYLDLDPV 1255
 DB 1201 GGAAPQHPPPAPSPAFDNLVYWDQPPERGA PPTSTKGTPTAENPEYLDLDPV 1255

RESULT 3
 AAB60167
 ID AAB60167 standard; Protein; 1255 AA.
 AC AAB60167;
 XX
 DT 03-APR-2001 (first entry)
 DE HER2 transgene plasmid construct encoded protein.
 XX
 KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
 XX antibody.
 OS Homo sapiens.
 XX Synthetic.
 XX
 PN WO200100244-A2.
 XX
 PD 04-JAN-2001.
 PF 23-JUN-2000; 2000WO-US17229.
 XX
 PR 25-JUN-1999; 99US-0141316.
 PR 16-MAR-2000; 2000US-0189844.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Erickson S, Schwall R;
 XX
 DR WPI; 2001-061962/07.
 DR N-PSDB; AAF24297.
 XX
 PT Treating tumors, particularly breast cancers, which overexpress an ErbB
 PT receptor and does not respond to an anti-ErbB antibody, comprises
 PT conjugating the antibody to a maytansinoid -

PS Example 3; Fig 4; 92pp; English.
 XX
 CC The present invention provides a method of treating cancer by
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
 CC particular, the antibody is directed against ErbB2 (also known as HER2
 CC and p185neu). The method is particularly useful in the treatment of
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.0%; Score 6627; DB 22; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 0;
 Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;
 QY 1 MELALCRWGLLALLPFGAASGYCTGDMKLRPAPETHLDMLRHLYGCGVYQGNL 60
 DB 1 MELALCRWGLLALLPFGAASGYCTGDMKLRPAPETHLDMLRHLYGCGVYQGNL 60
 QY 61 ELTYLPTNASLSFLQDIEVGQVYLIAHNOVQVBLQRLIVRGTLQFEDNYALAVDNG 120
 DB 61 ELTYLPTNASLSFLQDIEVGQVYLIAHNOVQVBLQRLIVRGTLQFEDNYALAVDNG 120
 QY 121 DPLNNTPTTASPGGLRELQRLSLTEILKGVYLQRRPOLCYOQTILMKDIFHKNNOLA 180
 DB 121 DPLNNTPTTASPGGLRELQRLSLTEILKGVYLQRRPOLCYOQTILMKDIFHKNNOLA 180
 QY 181 LTLIDTNRSRACHPSPCKSGRCWGESESDCQLTRVYACAGCARCKGRLPTDCHEQC 240
 DB 181 LTLIDTNRSRACHPSPCKSGRCWGESESDCQLTRVYACAGCARCKGRLPTDCHEQC 240
 QY 241 AAGCTGPKASDCLACHENHSGICEJHCPALVYNTDTFESMPNBEGRYTFGASCYTACP 300
 DB 241 AAGCTGPKASDCLACHENHSGICEJHCPALVYNTDTFESMPNBEGRYTFGASCYTACP 300
 QY 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGQRCCKSPCARVYCYGMQYIKANSKFCIT 360
 DB 301 YNYLSTDVSGCTLVCPRLHNOEYTAEDGQRCCKSPCARVYCYGMQYIKANSKFCIT 360
 QY 361 ELEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPDSL 420
 DB 361 IOEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPDSL 420
 QY 421 DLSVFQNLQVIGRILHNGAYSLTQGLISWLGSRSLRELSSGLALLHNNHLCFVHV 480
 DB 421 DLSVFQNLQVIGRILHNGAYSLTQGLISWLGSRSLRELSSGLALLHNNHLCFVHV 480
 QY 481 PMDQLFRNPHQALLHTANPEDECVGEGLACHQLCARHCWGPSTQCVNCSQFLRGQEC 540
 DB 481 PMDQLFRNPHQALLHTANPEDECVGEGLACHQLCARHCWGPSTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLOGLPREYVNAHGLPCHEBCOPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 DB 541 VEECRVLOGLPREYVNAHGLPCHEBCOPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 QY 601 PSGVXPDLASMPIMKFPDEEGACQCPINCTHSCVDLDKGGAPAEORASPLSIFNNFTV 660
 DB 601 PSGVXPDLASMPIMKFPDEEGACQCPINCTHSCVDLDKGGAPAEORASPLSIFNNFTV 660
 QY 661 SFMLRVKYSASHLEKRRQOKIRKTYMRLLQETELVEBLTPSGAMPNQAQWRILKETEL 720
 DB 661 ILLVVVLGVFGLIKRROQKIRKTYMRLLQETELVEBLTPSGAMPNQAQWRILKETEL 720
 QY 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780
 DB 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780
 QY 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEYDR 840
 DB 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEYDR 840
 QY 841 LVHRDLAARNVLYVSPNHVKTIDFGALRLDIDETEHAAOGKVPKIMMALESILRRFT 900
 DB 841 LVHRDLAARNVLYVSPNHVKTIDFGALRLDIDETEHAAOGKVPKIMMALESILRRFT 900

Db 841 LVHRDLAARVIVKSPHVKITDPLGLARLLDIDETEVHADGKVP2IKMMALESLRRRF 900
 Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGI PARSI PDLEKXGBLP PPICTIDVYIMVKCW 960
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLEKXGBLP PPICTIDVYIMVKCW 960
 Qy 961 IDSCRRPFELVSEFSRMRADPFRVIONEDLGPASPLDSTFFSYSLIEDDMGLVDA 1020
 Db 961 IDSCRRPFELVSEFSRMRADPFRVIONEDLGPASPLDSTFFSYSLIEDDMGLVDA 1020
 Qy 1021 EBYLVPOQGFCCPDPAFAGAGMYHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSE 1080
 Db 1021 EBYLVPOQGFCCPDPAFAGAGMYHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSE 1080
 Qy 1081 AGSDVFPDGLGMAKAKLQSLPTHDESP LORYSDDPVP LPSSTDGVVALTLTSPQPEYV 1140
 Db 1081 AGSDVFPDGLGMAKAKLQSLPTHDESP LORYSDDPVP LPSSTDGVVALTLTSPQPEYV 1140
 Qy 1141 NOPDVRPQPPSPREGPLPARPAGATLERAKTISPKNVGVKDVAFGAVENPEYLTQ 1200
 Db 1141 NOPDVRPQPPSPREGPLPARPAGATLERAKTISPKNVGVKDVAFGAVENPEYLTQ 1200
 Qy 1201 GGAAPQPPPPAPSPAFDNL YWDODPPERGAPSPSTFKGPTAENPEYGLDVPV 1255
 Db 1201 GGAAPQPPPPAPSPAFDNL YWDODPPERGAPSPSTFKGPTAENPEYGLDVPV 1255

RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 XX 23-APR-2002 (first entry)
 DE Human HER2 (ErbB2) polypeptide.
 XX
 XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumor; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytal disorder; hypothalamic disorder;
 KW glandular disorder; macropneal disorder; epithelial disorder;
 KW stromal disorder; blastococic disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 XX Homo sapiens.
 OS
 XX
 XX US2002001587-A1.
 PN
 XX 03-JAN-2002.
 PD
 XX
 XX 16-MAR-2001; 2001US-0811123.
 PF
 XX
 XX 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 XX (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIM/) SLIKOWSKI M.
 XX
 XX Erickson S, Schwall R, Slikowski M;
 XX
 XX WFI, 2002-163686/21.
 DR N-PSDB; ABK14058.
 DR
 XX Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 PS Example 3; Fig 7; 93pp; English.
 XX
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macropneal,
 CC epithelial, stromal, blastococic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 97.0%; Score 6627; DB 23; Length 1255;
 Best Local Similarity 97.4%; Pred. No. 0;
 Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

Qy 1 MELALCRWGLLALLPQAASTQVCTGDMKRLPASPEHLDRLHLYGCGVYQNL 60
 Db 1 MELALCRWGLLALLPQAASTQVCTGDMKRLPASPEHLDRLHLYGCGVYQNL 60
 Qy 61 ELTYPTNASIFLQDIEVQGYVLAHNOVROYQLQRLTYRGQLFEDYVALAVLNG 120
 Db 61 ELTYPTNASIFLQDIEVQGYVLAHNOVROYQLQRLTYRGQLFEDYVALAVLNG 120
 Qy 61 ELTYPTNASIFLQDIEVQGYVLAHNOVROYQLQRLTYRGQLFEDYVALAVLNG 120
 Db 61 ELTYPTNASIFLQDIEVQGYVLAHNOVROYQLQRLTYRGQLFEDYVALAVLNG 120
 Qy 121 DELNNTPTVTSFGLRELOLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNOLA 180
 Db 121 DELNNTPTVTSFGLRELOLRSLEILKGVLIQRNPOLCYODTILMKDIFHKNOLA 180
 Qy 181 LTLIDTNSRACHPSSPMCKSKRGSESSDCCSITRYVCGGRCRGPLPTDCHEQC 240
 Db 181 LTLIDTNSRACHPSSPMCKSKRGSESSDCCSITRYVCGGRCRGPLPTDCHEQC 240
 Qy 241 AAGCTGPRKSDCLACLFHFNHSGICELHPALVTVTDTFESMPNEGRYTGGASCVTPAC 300
 Db 241 AAGCTGPRKSDCLACLFHFNHSGICELHPALVTVTDTFESMPNEGRYTGGASCVTPAC 300
 Qy 301 YNYLSTDVGSCTLVCPFLHNOEVTADGTQRCCKSKPCARVYCYGIMQYIRANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPFLHNOEVTADGTQRCCKSKPCARVYCYGIMQYIRANSKFIGIT 360
 Qy 361 ELFEAGCKKIGSLAFIPESFPDPSASTAALQPOLQVFTLEITGYLYISAMPDSL 420
 Db 361 ELFEAGCKKIGSLAFIPESFPDPSASTAALQPOLQVFTLEITGYLYISAMPDSL 420
 Qy 421 DLVAFQNLQVIRGRILHNGAYSLTLQGLISMLGLRSRLREGSGALIHNTTHLCFVHTV 480
 Db 421 DLVAFQNLQVIRGRILHNGAYSLTLQGLISMLGLRSRLREGSGALIHNTTHLCFVHTV 480
 Qy 481 FMDOLFRRPHQALLHTANRPEDECVGEGSLACHQICARGHCHGPGPTQCVNCSQPLRGEC 540
 Db 481 FMDOLFRRPHQALLHTANRPEDECVGEGSLACHQICARGHCHGPGPTQCVNCSQPLRGEC 540
 Qy 541 VEECGVLOGLPREVYNARHCLPCHPECCQPNQSGVTCFGBADQCAACHYADPFCVARG 600
 Db 541 VEECGVLOGLPREVYNARHCLPCHPECCQPNQSGVTCFGBADQCAACHYADPFCVARG 600
 Qy 601 PSQVFPDLSYPIKFPDEBAGACPCPINTHSCVDLDDKCCPAEQRASPLTISVNFV 660
 Db 601 PSQVFPDLSYPIKFPDEBAGACPCPINTHSCVDLDDKCCPAEQRASPLTISVNFV 660
 Qy 661 SFWLRYPVYSASHLEKRGQKIRKVTMRRLLOETELVEPLTPSGAMPQAOARLKEHEL 720
 Db 661 SFWLRYPVYSASHLEKRGQKIRKVTMRRLLOETELVEPLTPSGAMPQAOARLKEHEL 720
 Qy 721 RKYVLGSGAGFYVYKGIWIPDGENVXIPVAKYLRNTSEKXKEILDEAYVAVAGVSP 780
 Db 721 RKYVLGSGAGFYVYKGIWIPDGENVXIPVAKYLRNTSEKXKEILDEAYVAVAGVSP 780
 Qy 781 YVSRLLGICLSTVQVLTQMLPYGCLLDHRENRGRIGSQDLNWCQIAKXGMYLEDVR 840

```

Db 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHREKRGSGSODLNMCMQIAKMSYLEDR 840
QY 841 LVHBDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESTLRRT 900
Db 841 LVHBDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESTLRRT 900
QY 901 HQSDWMSGVTVWELMTFGAKPYDGIPIAREIPDLLEGEGERLPOPICTIDVYIMVYCM 960
Db 901 HQSDWMSGVTVWELMTFGAKPYDGIPIAREIPDLLEGEGERLPOPICTIDVYIMVYCM 960
QY 961 IDSECRPFRELVESEFRMADPQRFVVIQNEDELGPASPLDSTFYSLLEDMDGLVDA 1020
Db 961 IDSECRPFRELVESEFRMADPQRFVVIQNEDELGPASPLDSTFYSLLEDMDGLVDA 1020
QY 1021 EEYLVPOGFCFPPAGAGGMVHRHRSSTRSGGDLTLGLPESEEARSPPLASEG 1080
Db 1021 EEYLVPOGFCFPPAGAGGMVHRHRSSTRSGGDLTLGLPESEEARSPPLASEG 1080
QY 1081 AGSDVPDGLDGMGAKGLOSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVPDGLDGMGAKGLOSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPVRPQPPSPRSGPLPAARPAATLEBAKTLSPGXGVYKDYAFAGAVENPEYLTPO 1200
Db 1141 NOPVRPQPPSPRSGPLPAARPAATLEBAKTLSPGXGVYKDYAFAGAVENPEYLTPO 1200
QY 1201 GAAPOPHPPPAFSPAFDNLVYWDQDPPERGAPSTPKGTPTAENPEYLGIDVY 1255
Db 1201 GAAPOPHPPPAFSPAFDNLVYWDQDPPERGAPSTPKGTPTAENPEYLGIDVY 1255

```

RESULT 5

AAW01111 standard; Protein; 1255 AA.

```

XX AC AAW01111;
XX DT 01-JAN-1997 (first entry)
XX DE HER-2/neu protein.
XX KM HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX KM breast cancer; ovary cancer; colon cancer; lung cancer;
XX KM prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 676..1255
XX FT /label= Intracellular domain
XX FT /note= "claimed domain, useful for immunisation"
XX FN W09630514-A1.
XX PD 03-OCT-1996.
XX PF 28-MAR-1996; 96WO-US01689.
XX PR 31-MAR-1995; 95US-0414417.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Cheever MA, Disis ML;
XX DR WPI: 1996-455361/45.
XX DR N-PSDB: AAT40739.
XX PT DNA encoding HER-2/neu poly(peptide(s)) - used for prevention or
XX PT treatment of malignancies with which the HER-2/neu oncogene is
XX PS associated
XX PS Claim 2; Page 56-61; 71pp; English.

```

```

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;

```

```

Query Match 96.9%; Score 6623; DB 17; Length 1255;
Best local Similarity 97.3%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

```

```

QY 1 MEIAALCRNGLLALIPGMASTVCTGDMKLRIPASPTHLDMLRHLVYCCQVQVNL 60
Db 1 MEIAALCRNGLLALIPGMASTVCTGDMKLRIPASPTHLDMLRHLVYCCQVQVNL 60
QY 61 ELTYLPTNASLSFLDIOEVQGVLIANQVQVPLQRLIRVGTQLFEDNALAVLNG 120
Db 61 ELTYLPTNASLSFLDIOEVQGVLIANQVQVPLQRLIRVGTQLFEDNALAVLNG 120
QY 121 DPANNTPTVPGASPGGLRELQSLTEILKGVLIQRPOLCYODTILMKDI FHKNNQLA 180
Db 121 DPANNTPTVPGASPGGLRELQSLTEILKGVLIQRPOLCYODTILMKDI FHKNNQLA 180
QY 181 LTLIDPNERACHPSPKSGRCWGESEDDQSILTRYCAGGACRCKPLPTDCCHQC 240
Db 181 LTLIDPNERACHPSPKSGRCWGESEDDQSILTRYCAGGACRCKPLPTDCCHQC 240
QY 241 AAGCTGPKSDCLACLFHNSGICELHCALVTYNTDFESMNPDEGRYTFGASCYACP 300
Db 241 AAGCTGPKSDCLACLFHNSGICELHCALVTYNTDFESMNPDEGRYTFGASCYACP 300
QY 301 YNYLSTDVGSCTLVCPDLNQBVTABDGTORCEKSKPCARVCYIGLMQYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPDLNQBVTABDGTORCEKSKPCARVCYIGLMQYIKANSKFIGIT 360
QY 361 ELPEAGCKKIFGSLAFIPSPFGDDPASNTAPLOPQLOVEFTEETGVLISAMPDLSL 420
Db 361 IOEFACCKKIFGSLAFIPSPFGDDPASNTAPLOPQLOVEFTEETGVLISAMPDLSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAVSLTLQGLISWLGRLSRLRELSGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAVSLTLQGLISWLGRLSRLRELSGLALIHNNHLCFVHTV 480
QY 481 PMDQLFRRNHQALLHTANPEDECVGEGLACHQLCARHGCMWPGPTCCMNGSOFJRGQC 540
Db 481 PMDQLFRRNHQALLHTANPEDECVGEGLACHQLCARHGCMWPGPTCCMNGSOFJRGQC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHECQFONGSVTCFGEADQCAVACHYKDPFCVARG 600
Db 541 VEECRVLOGLPREYVNAHCLPCHECQFONGSVTCFGEADQCAVACHYKDPFCVARG 600
QY 601 PSGVYRDLSTYMPIMKFPDEEGACQPCINCHSCVDLDDKGPAPQORASPLTSINNFV 660
Db 601 PSGVYRDLSTYMPIMKFPDEEGACQPCINCHSCVDLDDKGPAPQORASPLTSINNFV 660
QY 661 SFMLRVPKVSASHLEKROQKIRKTYMRLLQETLVEPLTPSGAMPNOAQRIKXETEL 720
Db 661 SFMLRVPKVSASHLEKROQKIRKTYMRLLQETLVEPLTPSGAMPNOAQRIKXETEL 720
QY 721 RYKRVLGSGAFGTVYKGIIPDGENVKIPVAIKVIRENTSPRANKIIDEAYVMAGVSP 780
Db 721 RYKRVLGSGAFGTVYKGIIPDGENVKIPVAIKVIRENTSPRANKIIDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHREKRGSGSODLNMCMQIAKMSYLEDR 840
Db 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHREKRGSGSODLNMCMQIAKMSYLEDR 840

```

QY 841 LVHRDLAARNVAVKSPNHVKTIDFGIARLDIDETEVHADGKVPKIMMALESTLRRT 900
 Db 841 LVHRDLAARNVAVKSPNHVKTIDFGIARLDIDETEVHADGKVPKIMMALESTLRRT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGPARE:PDLLKXGERLPQPPICITIDVYIMVCM 960
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGPARE:PDLLKXGERLPQPPICITIDVYIMVCM 960
 QY 961 IDSECRPRFRELVSFSFMADDFQFVYITONEDLGPSPLDSTYRSLLEDDMDLVDA 1020
 Db 961 IDSECRPRFRELVSFSFMADDFQFVYITONEDLGPSPLDSTYRSLLEDDMDLVDA 1020
 QY 1021 EEVLVPOGFCPPAPAGAGMWHRRSSSTRSGGGLTGLPSESEAPRSLAPSEG 1080
 Db 1021 EEVLVPOGFCPPAPAGAGMWHRRSSSTRSGGGLTGLPSESEAPRSLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPPQEV 1140
 Db 1081 AGSDVFDGDLGMAKGLQSLPTHDPSPLQRYSDPTVPLPSETDGYVAPLTCSPPQEV 1140
 QY 1141 NOPDVRPQPSRPSRGPAPARPAATLERAKTSPGKNGVYKVFARFAGAVENPEYLTPO 1200
 Db 1141 NOPDVRPQPSRPSRGPAPARPAATLERAKTSPGKNGVYKVFARFAGAVENPEYLTPO 1200
 QY 1201 GGAPQPHPPAPSPAFDNLTYMDQDPPERGAPSTFKGPTAENPEYLGIDVYV 1255
 Db 1201 GGAPQPHPPAPSPAFDNLTYMDQDPPERGAPSTFKGPTAENPEYLGIDVYV 1255

RESULT 6
 AAW92406
 ID AAW92406 standard; Protein: 1255 AA.

AAW92406;

DT 21-APR-1999 (first entry)

DE Human HER-2/neu oncogene protein.

KM HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 malignancy; treatment; tumour.

OS Homo sapiens.

FX Key Location/Qualifiers

FT Region 676..1255
 /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

XX 17-MAR-1993; 93US-0033644.

XX 12-AUG-1993; 93US-0106112.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

DR WPI, 1999-152835/13.

DR N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to

PT an HER-2/neu associated malignancy, particularly for treating or

XX preventing tumours

PS Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or recurrence.

XX Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 20; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALCRNGLLIALLPPGAASQVCTGTDMKRLPASBETHLDMRLHYOGGVQGNL 60
 Db 1 MELAALCRNGLLIALLPPGAASQVCTGTDMKRLPASBETHLDMRLHYOGGVQGNL 60
 QY 61 ELTYLPTNLSLFLQIOEVQGVYLAHQVROVPLQRLRYRGTOLEFENYALAVLDNG 120
 Db 61 ELTYLPTNLSLFLQIOEVQGVYLAHQVROVPLQRLRYRGTOLEFENYALAVLDNG 120
 QY 121 DPINNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
 Db 121 DPINNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
 QY 181 LTIIDNRSRACHPSCPKGSRGWESSESDQSLTRTYCAGGCACGKPLPTDCCHEOC 240
 Db 181 LTIIDNRSRACHPSCPKGSRGWESSESDQSLTRTYCAGGCACGKPLPTDCCHEOC 240
 QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTPESMNPGRATPFASCTYAC 300
 Db 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTPESMNPGRATPFASCTYAC 300
 QY 301 YNYLSTDVSGCTLVCPHNOEYTABDGTORCEKSPCARVCYGLGMOYIKANSKFTGIT 360
 Db 301 YNYLSTDVSGCTLVCPHNOEYTABDGTORCEKSPCARVCYGLGMOYIKANSKFTGIT 360
 QY 361 ELEFACCKKIFGSLAPSPGDPASNTAPLOPBOLOVFEETLGGVLYTSAMWDSLP 420
 Db 361 ELEFACCKKIFGSLAPSPGDPASNTAPLOPBOLOVFEETLGGVLYTSAMWDSLP 420
 QY 421 DLSVPONTQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGGLALIHNNHILCFVATV 480
 Db 421 DLSVPONTQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGGLALIHNNHILCFVATV 480
 QY 481 PMDQLFRNPHQALLHTANRPDECEVGEGLACHQLCARHGWGPGTQCVNCSQFLRGDEC 540
 Db 481 PMDQLFRNPHQALLHTANRPDECEVGEGLACHQLCARHGWGPGTQCVNCSQFLRGDEC 540
 QY 541 VEECRVLOGLPREYVVARHCLPCHPECCQPNQNSVTCFGEADQCVACAHYKOPPCVANC 600
 Db 541 VEECRVLOGLPREYVVARHCLPCHPECCQPNQNSVTCFGEADQCVACAHYKOPPCVANC 600
 QY 601 PSQVXPDLISYMPDKPEDEGACQPCPINCTHSVYLDKQCPAEGRASPLTISFNNFTV 660
 Db 601 PSQVXPDLISYMPDKPEDEGACQPCPINCTHSVYLDKQCPAEGRASPLTISFNNFTV 660
 QY 661 SFMLRPKYASASHLERROCKTRKXTMRLLLOETELVEZLTSGAMPNQAQRIIKETEL 720
 Db 661 SFMLRPKYASASHLERROCKTRKXTMRLLLOETELVEZLTSGAMPNQAQRIIKETEL 720
 QY 721 RKVKVLGSGAFGTYYKGIWIPDGENYKIPVAIKVLRNTPSPRANKEILDEAVYMGVGP 780
 Db 721 RKVKVLGSGAFGTYYKGIWIPDGENYKIPVAIKVLRNTPSPRANKEILDEAVYMGVGP 780
 QY 781 YVSRLLGLTSTVQLVQMLMPYGLLDHVRNRGRLSQDLINCMQIAKMSYLEDVR 840
 Db 781 YVSRLLGLTSTVQLVQMLMPYGLLDHVRNRGRLSQDLINCMQIAKMSYLEDVR 840
 QY 841 LVHRDLAARNVAVKSPNHVKTIDFGIARLDIDETEVHADGKVPKIMMALESTLRRT 900
 Db 841 LVHRDLAARNVAVKSPNHVKTIDFGIARLDIDETEVHADGKVPKIMMALESTLRRT 900

CC these neoplasias.
 XX Sequence 1255 AA;
 Query Match 96.9%; Score 6623; DB 21; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

CC 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 XX HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 QY 961 IDSECRPRRELVSSESMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020
 Db 961 IDSECRPRRELVSSESMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020
 QY 1021 EBYLVPOQGFCCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080
 Db 1021 EBYLVPOQGFCCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080
 QY 1081 AGSDVFDLDMGAAGKQSLPTHDPSLQRYSEDPYPLPSEEDGVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDLDMGAAGKQSLPTHDPSLQRYSEDPYPLPSEEDGVAPLTCSPQPEYV 1140
 QY 1141 NOPDVRFQPSPREGGLPAARPAAGATLEBAKTLSPGKGVKDYFAFGAAYENPEYLTPO 1200
 Db 1141 NOPDVRFQPSPREGGLPAARPAAGATLEBAKTLSPGKGVKDYFAFGAAYENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 7
 AAB21198 standard; protein, 1255 AA.
 ID AAB21198
 XX AAB21198;
 AC 12-JAN-2001 (first entry)
 DE Human HER-2/neu protein.
 KW Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.
 OS Homo sapiens.
 XX MO200044899-A1.
 XX 03-AUG-2000.
 XX 28-JAN-2000; 2000MO-US02164.
 XX 29-JAN-1999; 99US-0117976.
 XX (CORI-) CORIXA CORP.
 XX (SMIK) SMITHKLINE BEECHAM.
 PI Cheever MA, Gheyssen D;
 DR WPI: 2000-505976/45.
 DR N-PSDB: AAA89736.
 XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX Claim 52; Fig 7; 128pp; English.
 CC The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 XX HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 QY 961 IDSECRPRRELVSSESMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020
 Db 961 IDSECRPRRELVSSESMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020
 QY 1021 EBYLVPOQGFCCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080
 Db 1021 EBYLVPOQGFCCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080
 QY 1081 AGSDVFDLDMGAAGKQSLPTHDPSLQRYSEDPYPLPSEEDGVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDLDMGAAGKQSLPTHDPSLQRYSEDPYPLPSEEDGVAPLTCSPQPEYV 1140
 QY 1141 NOPDVRFQPSPREGGLPAARPAAGATLEBAKTLSPGKGVKDYFAFGAAYENPEYLTPO 1200
 Db 1141 NOPDVRFQPSPREGGLPAARPAAGATLEBAKTLSPGKGVKDYFAFGAAYENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255

CC 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 XX HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960
 QY 961 IDSECRPRRELVSSESMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020
 Db 961 IDSECRPRRELVSSESMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020
 QY 1021 EBYLVPOQGFCCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080
 Db 1021 EBYLVPOQGFCCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080
 QY 1081 AGSDVFDLDMGAAGKQSLPTHDPSLQRYSEDPYPLPSEEDGVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDLDMGAAGKQSLPTHDPSLQRYSEDPYPLPSEEDGVAPLTCSPQPEYV 1140
 QY 1141 NOPDVRFQPSPREGGLPAARPAAGATLEBAKTLSPGKGVKDYFAFGAAYENPEYLTPO 1200
 Db 1141 NOPDVRFQPSPREGGLPAARPAAGATLEBAKTLSPGKGVKDYFAFGAAYENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255

```

Db      961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGPRASPLDSTFYSRLLEDDDDGDLVDA 1020
Qy      1021 EBYLVPOQGFCCPDPAFAGAGMVAHRRSSSTRSGGDLTLGLEPSEEAARSPLEPSEG 1080
Db      1021 EBYLVPOQGFCCPDPAFAGAGMVAHRRSSSTRSGGDLTLGLEPSEEAARSPLEPSEG 1080
Qy      1081 AGSDVFPQGDLMGAAKGLQSLPTHDPSPLOQYSDPVPVLPSETDGVVAALTGSPQPEYV 1140
Db      1081 AGSDVFPQGDLMGAAKGLQSLPTHDPSPLOQYSDPVPVLPSETDGVVAALTGSPQPEYV 1140
Qy      1141 NQPDVPRQPSPREBPLPAARPAATLERAKTLSPGKNGVKDVFAGAVENPEYLTPO 1200
Db      1141 NQPDVPRQPSPREBPLPAARPAATLERAKTLSPGKNGVKDVFAGAVENPEYLTPO 1200
Qy      1201 GGAPQPHPPAPSPADNLYYNDQDPPERGAPSTPKGPTLAEPEYLGLDVAV 1255
Db      1201 GGAPQPHPPAPSPADNLYYNDQDPPERGAPSTPKGPTLAEPEYLGLDVAV 1255

```

RESULT 8

```

ID      AAY84780 standard; Protein; 1255 AA.
AC      AAY84780;
DT      08-AUG-2000 (first entry)
XX

```

DE Amino acid sequence of the SPLICE erbB-2 receptor protein.

KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;
 tumor cell proliferation; tissue degeneration; arthropathy;
 bone resorption; inflammatory disease; degenerative disorder;
 wound healing.

OS Homo sapiens.

PN WC000020579-A1.

PD 13-APR-2000.

PF 01-OCT-1999; 99WO-CA00912.

PR 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

PA Muller WJ, Siegel PM;

PI WPI, 2000-303768/26.

DR N-PSDB; AAA14812.

XX Nucleic acid encoding an erbB 2 receptor protein designated SPLICE

PT erbB-2, inhibitors of the protein are useful for treatment of cancer -

PS Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbB-2 receptor protein. The
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are
 CC conserved cysteine residues, compared to the unspliced protein. The
 CC erbB-2 polynucleotide is used to construct probes for detecting
 CC disorders of cell transformation such as cancer. Antibodies to the
 CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
 CC (e.g., antisense oligonucleotides) which inhibit the expression of
 CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
 CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
 CC for treating conditions involving damaged cells including conditions
 CC in which degeneration of tissue occurs, such as arthropathy, bone
 CC resorption, inflammatory diseases, degenerative disorders of the
 CC central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 21; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

```

Qy      1 MELIACRMGILLALLPQASTVOYCTGDMRLRLPASFTLDMRLHYOGQVVOGSL 60
Db      1 MELIACRMGILLALLPQASTVOYCTGDMRLRLPASFTLDMRLHYOGQVVOGSL 60
Qy      61 ELTYLPTNASTSFLQDIDQVGYVLIANQVQVPLQRLIRYGTQLFEDNALVALVDNG 120
Db      61 ELTYLPTNASTSFLQDIDQVGYVLIANQVQVPLQRLIRYGTQLFEDNALVALVDNG 120
Qy      121 DELNNTPTTSGSPGSLRELQRLSTLEIKGVLIGRNPQLCYOITLWKDIFHKNOQA 180
Db      121 DELNNTPTTSGSPGSLRELQRLSTLEIKGVLIGRNPQLCYOITLWKDIFHKNOQA 180
Qy      121 DELNNTPTTSGSPGSLRELQRLSTLEIKGVLIGRNPQLCYOITLWKDIFHKNOQA 180
Db      121 DELNNTPTTSGSPGSLRELQRLSTLEIKGVLIGRNPQLCYOITLWKDIFHKNOQA 180
Qy      181 LTLIDNRRACHCPSPMKSGRSGESEDQSLTRTVACAGCARCKGPLETDCHEQC 240
Db      181 LTLIDNRRACHCPSPMKSGRSGESEDQSLTRTVACAGCARCKGPLETDCHEQC 240
Qy      241 AAGCTGPKASDCLACLFENHSGICEAHGPAVTYNTDFESMPNEGRYTFGASCVTACP 300
Db      241 AAGCTGPKASDCLACLFENHSGICEAHGPAVTYNTDFESMPNEGRYTFGASCVTACP 300
Qy      301 VNYLSTDVSSCTLVGRHLNQEYTAEDGTQRCCKSPCARVCYGLGMYIKANSKFTGTT 360
Db      301 VNYLSTDVSSCTLVGRHLNQEYTAEDGTQRCCKSPCARVCYGLGMYIKANSKFTGTT 360
Qy      361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFEITGTYLSAMPDSL 420
Db      361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFEITGTYLSAMPDSL 420
Qy      421 DLIVQNTQVITRGLIANGAYSLTQSGISMLGRLSRLGSGIALIHHNTHLCFVATV 480
Db      421 DLIVQNTQVITRGLIANGAYSLTQSGISMLGRLSRLGSGIALIHHNTHLCFVATV 480
Qy      481 FWDQLEFRPHOALLTANRPEDECEGEGLACHOC-CARGHCMPGPTQCVNCSQFLRGEC 540
Db      481 FWDQLEFRPHOALLTANRPEDECEGEGLACHOC-CARGHCMPGPTQCVNCSQFLRGEC 540
Qy      541 VEECRVLTQGLPREVYNARHCLPCHBECCPONGSVTCFPEADQCAVACHYKDPFCVARG 600
Db      541 VEECRVLTQGLPREVYNARHCLPCHBECCPONGSVTCFPEADQCAVACHYKDPFCVARG 600
Qy      601 PSQVPELSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCPAEQRASPLTSLFNFTV 660
Db      601 PSQVPELSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCPAEQRASPLTSLFNFTV 660
Qy      661 SFWLKVPKVSASHLEKRRQOKIRKYTMRLLOETELVEPLTPSGAMPQAOARILKETEL 720
Db      661 SFWLKVPKVSASHLEKRRQOKIRKYTMRLLOETELVEPLTPSGAMPQAOARILKETEL 720
Qy      721 RRYKYLGSAGATGYKGIWIPDGENVKIPVAIKYLRENTSPKANKEILDEAYVAVAGSP 780
Db      721 RRYKYLGSAGATGYKGIWIPDGENVKIPVAIKYLRENTSPKANKEILDEAYVAVAGSP 780
Qy      781 YVSRLLGICLTSTQVLTQULMPYGCILDHRENGRGLSGQILLMWCMQIAGMSYLEDVR 840
Db      781 YVSRLLGICLTSTQVLTQULMPYGCILDHRENGRGLSGQILLMWCMQIAGMSYLEDVR 840
Qy      841 LVHRLDLAARNLVSPNHVKITDFGLARLIDIDETEHADGKVPIMMALESILRRFT 900
Db      841 LVHRLDLAARNLVSPNHVKITDFGLARLIDIDETEHADGKVPIMMALESILRRFT 900
Qy      901 HQSVWVSQVTVWELMTFGAKPYDGIPIAREIPDLLEGERLPQPICTIDVYIMVCM 960
Db      901 HQSVWVSQVTVWELMTFGAKPYDGIPIAREIPDLLEGERLPQPICTIDVYIMVCM 960
Qy      961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGPRASPLDSTFYSRLLEDDDDGDLVDA 1020
Db      961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDEGPRASPLDSTFYSRLLEDDDDGDLVDA 1020
Qy      1021 EBYLVPOQGFCCPDPAFAGAGMVAHRRSSSTRSGGDLTLGLEPSEEAARSPLEPSEG 1080

```



```

DB 1021 EEYIVPQGFCCPBPAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1060
QY 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
DB 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPSPPREGPLPAARPAAGATLEBAKTLSPGKNGVYKVFARFGAVENPEYLTPQ 1200
DB 1141 NOPDVRPQPSPPREGPLPAARPAAGATLEBAKTLSPGKNGVYKVFARFGAVENPEYLTPQ 1200
QY 1201 GGAAQPPHPAPFSPAFDNLVYWDQDPERGAPPTKGTPTAENPEYLGLDVPV 1255
DB 1201 GGAAQPPHPAPFSPAFDNLVYWDQDPERGAPPTKGTPTAENPEYLGLDVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
AC AAB85458;
XX
XX 25-SEP-2001 (first entry)
DE Human HER-2/neu protein.
XX Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KM oncogene; cancer; cytostatic; vaccine; p185; C-erbB2.
XX
XX Homo sapiens.
OS
XX WO200153463-A2.
PN
XX 26-JUL-2001.
PD
XX 19-JAN-2001; 2001WO-US01850.
FE
XX 21-JAN-2000; 2000US-0177545.
PR
XX (CORI-) CORIXA CORP.
XX
XX Cheever MA, Hand-Zimmermann S;
PI
XX MPI: 2001-476112/51.
DR N-PSDB; AAH23392.
DR
XX New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
XX The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or C-erbB2).
XX
XX Sequence 1255 AA;

Query March 96.9%; Score 6623; DB 22; Length 1255;
Best local similarity 97.3%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALGRMGLLLLPAGASTVCTGTDKMLPASPEHLDMLHLYOGQVYVGNL 60
DB 1 MELAALGRMGLLLLPAGASTVCTGTDKMLPASPEHLDMLHLYOGQVYVGNL 60
QY 61 ELTYLPTNASLFLQDIDQVGVYVLIANQVRQVPLQRLRVGTQLFEDNYVALAVLNDG 120

DB 61 ELTYLPTNASLFLQDIDQVGVYVLIANQVRQVPLQRLRVGTQLFEDNYVALAVLNDG 120
QY 121 DPLANTTPTTGASPGGLBELQSLTELKGVYLQRPOLCYOFTILMKOIFHNKOLA 180
DB 121 DPLANTTPTTGASPGGLBELQSLTELKGVYLQRPOLCYOFTILMKOIFHNKOLA 180
QY 181 LTLITNRSRACHPSPCKSGRCWGESSEDCQSLTRFYVAGGACRCARCKGAPLTDCCHEQC 240
DB 181 LTLITNRSRACHPSPCKSGRCWGESSEDCQSLTRFYVAGGACRCARCKGAPLTDCCHEQC 240
QY 241 AAGCTGPKXSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
DB 241 AAGCTGPKXSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
QY 301 YNYLSTDPVSCCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
DB 301 YNYLSTDPVSCCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360
QY 361 ELEFAGCKKIFGSLAFLESPDGPASNTAPLQPEOLOVFETLEITGYLYISAMPDSL 420
DB 361 ELEFAGCKKIFGSLAFLESPDGPASNTAPLQPEOLOVFETLEITGYLYISAMPDSL 420
QY 421 DLSVTQNLQVIRGRIILHNGAVSLTQGLISWLGRLSRLELGSGLALHNTHLCPVHTV 480
DB 421 DLSVTQNLQVIRGRIILHNGAVSLTQGLISWLGRLSRLELGSGLALHNTHLCPVHTV 480
QY 481 PMDQLFRNPHQALHTANRPEDECGEGLAQHLCARHGWPGFTQCVCNCSQFLRGQEC 540
DB 481 PMDQLFRNPHQALHTANRPEDECGEGLAQHLCARHGWPGFTQCVCNCSQFLRGQEC 540
QY 541 VEECRVQLGLPREYVYNAHCLPCHPECPONGSVTCFGEADQVACAHYKDPFCVARC 600
DB 541 VEECRVQLGLPREYVYNAHCLPCHPECPONGSVTCFGEADQVACAHYKDPFCVARC 600
QY 601 PSGVPRDLSYMPIMKPEDEGACQPCPINCCHSCVDLDKGPABQORASPLTISAVVG 660
DB 601 PSGVPRDLSYMPIMKPEDEGACQPCPINCCHSCVDLDKGPABQORASPLTISAVVG 660
QY 661 SFMLRVPKVSASHLEKRRQOKIRKTYMRLLOETELVBLTPSGAMPNQAQMRILKETEL 720
DB 661 SFMLRVPKVSASHLEKRRQOKIRKTYMRLLOETELVBLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFGTGKGIWIPDGEVYKIPVAIKVIRENTSPANKEILDEAVYMGVGP 780
DB 721 RKVKVLGSGAFGTGKGIWIPDGEVYKIPVAIKVIRENTSPANKEILDEAVYMGVGP 780
QY 781 YVSRLLGICLTSTQVLTQMLPYGCLLDHVENRGRGLSGDILNKCQIAKMSYLEYVR 840
DB 781 YVSRLLGICLTSTQVLTQMLPYGCLLDHVENRGRGLSGDILNKCQIAKMSYLEYVR 840
QY 841 LVHRDIAANVIVKSPNHVKITDGLARLLDIDETRYHADGKVPDKMALESILRRFT 900
DB 841 LVHRDIAANVIVKSPNHVKITDGLARLLDIDETRYHADGKVPDKMALESILRRFT 900
QY 901 HOSDWSYGVYTWELMTFEGAKPYDGIIPAREIDLEKGRILPOPICTIDYMIWVKCM 960
DB 901 HOSDWSYGVYTWELMTFEGAKPYDGIIPAREIDLEKGRILPOPICTIDYMIWVKCM 960
QY 961 IDSECRPRELSESRMARDPORFVYVIONEDLPASPFLSTFYRSLLEDDEDDMDLYDA 1020
DB 961 IDSECRPRELSESRMARDPORFVYVIONEDLPASPFLSTFYRSLLEDDEDDMDLYDA 1020
QY 1021 EEYIVPQGFCCPBPAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
DB 1021 EEYIVPQGFCCPBPAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
QY 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
DB 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPSPPREGPLPAARPAAGATLEBAKTLSGKNGVYKVFARFGAVENPEYLTPQ 1200
DB 1141 NOPDVRPQPSPPREGPLPAARPAAGATLEBAKTLSGKNGVYKVFARFGAVENPEYLTPQ 1200

```


Db 1141 NQDVRPQPSREGPLPAARPAATLERPKTSLPGKGVKQVFAFGAVENPEYLTPO 1200
 QY 1201 GGAAQPHPPAPSPADNLYYWDOPPERCAPSTKGTPTANPEYLDVPU 1255
 Db 1201 GGAAQPHPPAPSPADNLYYWDOPPERCAPSTKGTPTANPEYLDVPU 1255

RESULT 10
 AAG88267
 ID AAG88267 standard; Protein; 1255 AA.
 AC AAG88267;
 DT 11-SEP-2001 (first entry)
 DE HER2/neu amino acid sequence.
 KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KM tumour response; vaccine; cancer; cytotoxic; immunostimulant;
 OS Homo sapiens.
 PN MO200141787-A1.
 PD 14-JUN-2001.
 PF 11-DEC-2000; 2000WO-US33591.
 PR 10-DEC-1999; 99US-0458299.
 PA (EPIM-) EPIMUNE INC.
 PI Fikes J, Sette A, Sidney J, Southwood S, Cheesnut R, Celis E;
 PI Keogh E;
 DR WP1; 2001-374995/39.
 PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer -
 PS Disclosure; Page 15; 1999p; English.

The present invention describes isolated prepared HER2/neu epitopes (I).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (I), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (1); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic
 CC and immunostimulant activities, and can be used in vaccines. (II), (III)
 CC and (III) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (I) and (II) are useful for
 CC monitoring or evaluating an immune response to a tumour-associated
 CC antigen when incubated with a T lymphocyte sample form a patient and
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumour
 CC vaccines provides the opportunity to combine epitopes derived from
 CC multiple tumour-associated molecules addressing the problem of tumour-
 CC tumour variability and reducing the likelihood of tumour escape due to
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
 CC the exemplification of the present invention.

Sequence 1255 AA:
 Query Match 96.9%; Score 6623; DB 22; Length 1255;

Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPQASTQVCTGDMRLRPASBETHLDMRLHYQCCQVQVNL 60
 Db 1 MELAALCRWGLLALLPQASTQVCTGDMRLRPASBETHLDMRLHYQCCQVQVNL 60

QY 61 ELTYLPTNASLSFLDIOEVQGVLIANQVRQVPLQRIYRGTLQFEDNALAVLNG 120
 Db 61 ELTYLPTNASLSFLDIOEVQGVLIANQVRQVPLQRIYRGTLQFEDNALAVLNG 120

QY 121 DELNNTTPTVGTASPGGLRELQRLSTELIKGVLIQBNPOLCYOTILMKDIFHNQOLA 180
 Db 121 DELNNTTPTVGTASPGGLRELQRLSTELIKGVLIQBNPOLCYOTILMKDIFHNQOLA 180

QY 181 LTLIDTNRBRACHPCSPMCKGSRGWESSEDCQSLTRTVACAGCARCKGPLPTDCHEQC 240
 Db 181 LTLIDTNRBRACHPCSPMCKGSRGWESSEDCQSLTRTVACAGCARCKGPLPTDCHEQC 240

QY 241 AAGCTGPRKSDCLALCFPHNSGICELHCPALVTYNTDTPESMPNREGRTFGASCYTACP 300
 Db 241 AAGCTGPRKSDCLALCFPHNSGICELHCPALVTYNTDTPESMPNREGRTFGASCYTACP 300

QY 301 VNYLSTDVSSCTLVCPPLHNOEYTAEDGTORCEKSCPCARVCYGLMOTIKANSKFIGIT 360
 Db 301 VNYLSTDVSSCTLVCPPLHNOEYTAEDGTORCEKSCPCARVCYGLMOTIKANSKFIGIT 360

QY 361 ELTFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIYISAMPDILP 420
 Db 361 ELTFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIYISAMPDILP 420

QY 421 DLVSFQNIQVTRGRILHNGAYSLTLOGIGISWLGIRSLRELGSGLALHNHNLGCFVHTV 480
 Db 421 DLVSFQNIQVTRGRILHNGAYSLTLOGIGISWLGIRSLRELGSGLALHNHNLGCFVHTV 480

QY 481 PWDQLFRPHQALLHTANRPEDECVGEGLACHQLCARGHCWPGPTQCVNCSQFRLGQEC 540
 Db 481 PWDQLFRPHQALLHTANRPEDECVGEGLACHQLCARGHCWPGPTQCVNCSQFRLGQEC 540

QY 541 VEECVLQGLPREYVNAHCLPCHRECQPNQSVTCFEPDQCVACHYKDPFCVARC 600
 Db 541 VEECVLQGLPREYVNAHCLPCHRECQPNQSVTCFEPDQCVACHYKDPFCVARC 600

QY 601 PSQVPRDLSYPIWPKFPEBEGACQPCINCHSCVYDLDKGPAPORASPLTISANNVT 660
 Db 601 PSQVPRDLSYPIWPKFPEBEGACQPCINCHSCVYDLDKGPAPORASPLTISANNVT 660

QY 661 SFWRVPRVSAHLEKRRQKIRKYMRLQETELVLPITPSGAMPNQAQWRIKETEL 720
 Db 661 SFWRVPRVSAHLEKRRQKIRKYMRLQETELVLPITPSGAMPNQAQWRIKETEL 720

QY 721 RKVKYLSGAGATYKGIWIPDGEVVKIPVAILKRLRENTSKANKELIDEAYVWAGVSP 780
 Db 721 RKVKYLSGAGATYKGIWIPDGEVVKIPVAILKRLRENTSKANKELIDEAYVWAGVSP 780

QY 841 LVHRDLARNTLVSPNNVKTITDGLARLIDITETVADGKVPKIMALESILRRFT 900
 Db 841 LVHRDLARNTLVSPNNVKTITDGLARLIDITETVADGKVPKIMALESILRRFT 900

QY 901 HOSDWSAGVYVMEVMTFGAKYUGIPAREIPDLLEGGERLPORPCTIDVYMWVCMW 960
 Db 901 HOSDWSAGVYVMEVMTFGAKYUGIPAREIPDLLEGGERLPORPCTIDVYMWVCMW 960

QY 961 IDSECRPRFRELVSFERRMADPQRFVYIQNEDLGAPSLDSTFYRSLLEDDEMDGLVDA 1020
 Db 961 IDSECRPRFRELVSFERRMADPQRFVYIQNEDLGAPSLDSTFYRSLLEDDEMDGLVDA 1020

QY 1021 EBYLVPOGCFPCPPAPAGAGCMVHNRSSSTRGGDLTLGLEPSEEARPSGLASEG 1080

Db 1021 EEVLVPQGFCCPDPAAGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFPDGLMGAAKGIQSLPTHDPSPLQRYSEDPTVPLPSETDGYAAPLTGSPQPEYV 1140
Db 1081 AGSDVFPDGLMGAAKGIQSLPTHDPSPLQRYSEDPTVPLPSETDGYAAPLTGSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVYKDVFAFGAENBEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVYKDVFAFGAENBEYLTPO 1200
Qy 1201 GGAAFPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVYKDVFAFGAENBEYLTPO 1200
Db 1201 GGAAFPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVYKDVFAFGAENBEYLTPO 1200

RESULT 11

AAE24067
ID AAE24067 standard; Protein; 1255 AA.

AAE24067;
AC AAE24067;

DT 23-SEP-2002 (first entry)
XX

DE Human Her-2 protein.
XX

KM Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KM hyperproliferative disorder; prophylaxis; inflammation; antisense;
KM tumour; gene therapy; phosphorothioate backbone.
XX

XX Homo sapiens.
XX

XX WO200222636-A1.
XX

XX 21-MAR-2002.
XX

XX 12-SEP-2001; 2001MO-US28572.
XX

XX 15-SEP-2000; 2000US-0663834.
XX

XX (ISIS-) ISIS PHARM INC.
XX

XX Bennett CF, Cowser LM;
XX

XX WPI; 2002-471192/50.
XX

XX N-ESDB; AAD38904.
XX

XX Novel antisense oligonucleotide which modulates the expression of Human
XX Epidermal Growth Factor receptor, Her2, is useful for treating tumors
XX inflammation or to prevent infection in humans -
XX

XX Example 13; Page 95-107; 116pp; English.
XX

XX The invention relates to antisense compounds targeted to a nucleic
XX acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
XX that specifically hybridizes with and inhibits the expression of Her2.
XX Antisense compounds of the invention are used for treating diseases or
XX conditions associated with Her2 such as hyperproliferative disorders
XX e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
XX neutral or cardiac cancer. They are also useful prophylactically e.g.
XX to prevent or delay infection, inflammation and tumour formation. The
XX invention is also used in gene therapy. The present sequence is human
XX Her-2 protein.
XX

XX Sequence 1255 AA;
XX

XX Query Match 96.9%; Score 6623; DB 23; Length 1255;
XX Best Local Similarity 97.3%; Pred. No. 0;
XX Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

Qy 1 METALICWGLLALLPFGAASSTVCTGDMKRLPASPETHLMDLRLYLGGCCVAGNL 60
Db 1 METALICWGLLALLPFGAASSTVCTGDMKRLPASPETHLMDLRLYLGGCCVAGNL 60

Qy 1 METALICWGLLALLPFGAASSTVCTGDMKRLPASPETHLMDLRLYLGGCCVAGNL 60
Db 1 METALICWGLLALLPFGAASSTVCTGDMKRLPASPETHLMDLRLYLGGCCVAGNL 60

Qy 61 ELTYLPTNASLFLQDIOEVQGVLLAHNOVQVPLQRLRIVRGTQLFEDNALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEVQGVLLAHNOVQVPLQRLRIVRGTQLFEDNALAVLDNG 120
Qy 121 DPLNNTTPTVAGSPGGLRELQRLSLTEILKGVLLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db 121 DPLNNTTPTVAGSPGGLRELQRLSLTEILKGVLLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy 181 LTLIDNRSRACHPCSPMCKGRSMWSESESDCSLRTTACAGGACRCKPLPTDCHEOC 240
Db 181 LTLIDNRSRACHPCSPMCKGRSMWSESESDCSLRTTACAGGACRCKPLPTDCHEOC 240
Qy 241 AAGCTPKASDCLACHFNHSGICELHCPALVTYNTDTESPMNEGRYTFGASCTYACP 300
Db 241 AAGCTPKASDCLACHFNHSGICELHCPALVTYNTDTESPMNEGRYTFGASCTYACP 300
Qy 301 YNLTSTDVSCCTLVCEPHNQEVTAEDGTQRCCKSPCARVCYGLMEHLREYRAVTSAN 360
Db 301 YNLTSTDVSCCTLVCEPHNQEVTAEDGTQRCCKSPCARVCYGLMEHLREYRAVTSAN 360
Qy 361 ELFPACCKIFGSLAFLPESFDGDPASNTAPLOPBOLOVFETLEITGYLYISAMPDHP 420
Db 361 IOFPACCKIFGSLAFLPESFDGDPASNTAPLOPBOLOVFETLEITGYLYISAMPDHP 420
Qy 421 DLSVFNOLQYIRGRILHNGAYSLTLOGDISWGLRLSFLSGSLALIHNTLCHVHTV 480
Db 421 DLSVFNOLQYIRGRILHNGAYSLTLOGDISWGLRLSFLSGSLALIHNTLCHVHTV 480
Qy 481 PMDQLFRNPQALHNTANPEDECEVEGLACHOLCRGCHWGPGPTQVNCOSQFLGQEC 540
Db 481 PMDQLFRNPQALHNTANPEDECEVEGLACHOLCRGCHWGPGPTQVNCOSQFLGQEC 540
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECCQONGSVTCFGEADQCVACAHYKDPFCVAC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPECCQONGSVTCFGEADQCVACAHYKDPFCVAC 600
Qy 601 PSQVAPDLSTMPYMKPDEEGACQPCPINCTHSCVLDKDGCAEGRASPLTIFNNFTV 660
Db 601 PSQVAPDLSTMPYMKPDEEGACQPCPINCTHSCVLDKDGCAEGRASPLTIFNNFTV 660
Qy 661 SFMLRVPKVASHLERROOKIRKYMRLLOTELEVEPLTPSGAMPNOAMILKETEL 720
Db 661 SFMLRVPKVASHLERROOKIRKYMRLLOTELEVEPLTPSGAMPNOAMILKETEL 720
Qy 721 RYKVKVLSGAGFTVYKGIWIPDGENYKIPAIKVLRENTSPKANKELDEAYVAGVSP 780
Db 721 RYKVKVLSGAGFTVYKGIWIPDGENYKIPAIKVLRENTSPKANKELDEAYVAGVSP 780
Qy 781 YVSRLLGICLSTVQCLVTOQLMPYGCLLDHVRENRGLSGODLNMCMQIAKMSYLEDVR 840
Db 781 YVSRLLGICLSTVQCLVTOQLMPYGCLLDHVRENRGLSGODLNMCMQIAKMSYLEDVR 840
Qy 841 LVHRDLAARVNLVKSNNHYKITDFGLARLLDIDETEHADGKVPITKMALESILRRFT 900
Db 841 LVHRDLAARVNLVKSNNHYKITDFGLARLLDIDETEHADGKVPITKMALESILRRFT 900
Qy 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPLLEKGERLPPPICTIDVYIMWKCM 960
Db 901 HOSDWSYGVTTWELMTFGAKPYDGIIPAEIPLLEKGERLPPPICTIDVYIMWKCM 960
Qy 961 IDECEPRPRELISEFSRMAARDPQRVVIVQNEDELGPASPLDSFTFYSLEDDMGDLVDA 1020
Db 961 IDECEPRPRELISEFSRMAARDPQRVVIVQNEDELGPASPLDSFTFYSLEDDMGDLVDA 1020
Qy 1021 EEVLVPQGFCCPDPAAGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Db 1021 EEVLVPQGFCCPDPAAGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFPDGLMGAAKGIQSLPTHDPSPLQRYSEDPTVPLPSETDGYAAPLTGSPQPEYV 1140
Db 1081 AGSDVFPDGLMGAAKGIQSLPTHDPSPLQRYSEDPTVPLPSETDGYAAPLTGSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVYKDVFAFGAENBEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVYKDVFAFGAENBEYLTPO 1200

DB 1141 NQDVRPQPPSPRSGPLPAAPAGTLEPRKTLSPGKGVKDVFAFGAVENPEYLTPQ 1200
 QY 1201 GGAAQPPHPPAFSPAFDNLVYWDODPPERGAPSTPKGPTAENPEYLGLDVPV 1255
 DB 1201 GGAAQPPHPPAFSPAFDNLVYWDODPPERGAPSTPKGPTAENPEYLGLDVPV 1255

RESULT 12
 AAE20479 standard; Protein; 1255 AA.
 ID AAE20479
 AC AAE20479
 DT 01-JUL-2002 (first entry)
 DE Human Her-2/neu protein.
 KM Human; Her-2/Neu protein; immune response; gene therapy; breast cancer; human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 1021..1030
 FT /note= "Naturally processed HLA-B44-restricted epitope"
 XX WO200214503-A2.
 XX 21-FEB-2002.
 XX 14-AUG-2001; 2001WO-US41733.
 XX 14-AUG-2000; 2000US-225153P.
 XX 28-SEP-2000; 2000US-236428P.
 XX 21-FEB-2001; 2001US-270520P.
 XX (CORI-) CORIXA CORP.
 PI Hand-Zimmermann S, Cheever MA, Foy TM, Lodes MT, Kalos MD;
 PI McNeill PD, Vedvick TS;
 DR WPI; 2002-280758/32.
 DR N-PSDB; AAD32743.
 XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
 PT prevention and diagnosis of cancer, preferably breast cancer
 PS Disclosure; Page 114-117, 129pp; English.
 XX The invention relates to an isolated Her-2/Neu polypeptide composition
 CC effective for eliciting an immune response. The invention is useful for
 CC eliciting an immune response in a patient where the patient is human
 CC leukocyte antigen (HLA) B44 positive or is affected with breast cancer.
 CC The composition is useful for the therapy and diagnosis of cancer.
 CC Preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
 CC and other compositions for the diagnosis, prevention and treatment of
 CC human malignancies, for stimulating and/or expanding T cells specific for
 CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
 CC patient. The invention is useful for stimulating a T cell response in a
 CC human patient, as probe or primer for nucleic acid hybridisation, to
 CC selectively form duplex molecules with complementary stretches of the
 CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
 CC length gene from a suitable library, and to direct expression of a
 CC polypeptide in appropriate host cells. The composition is useful in
 CC prophylactic or therapeutic applications and for the treatment of cancer,
 CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
 CC associated malignancies. The invention is useful in gene therapy. The
 CC present sequence is human Her-2/neu protein.
 SQ Sequence 1255 AA;

Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAAACRMGLLALLPFGAASQVCTGDMKRLPASPETHLDMRHLVYGGQVYQGNL 60
 DB 1 MELAAACRMGLLALLPFGAASQVCTGDMKRLPASPETHLDMRHLVYGGQVYQGNL 60

QY 61 ELTYPTNALSLEFODIOEVQGVYLAHNVQVPLQVRJRVGTQLFEDNVALALVLDNG 120
 DB 61 ELTYPTNALSLEFODIOEVQGVYLAHNVQVPLQVRJRVGTQLFEDNVALALVLDNG 120

QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNQLCYQDTLLKQDFHKNQOLA 180
 DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNQLCYQDTLLKQDFHKNQOLA 180

QY 181 LTLIDNRRSRAHPCSPMCKSGRCWSESSEDCOSLTRTCAGAGCARCKPLPDDCCEQC 240
 DB 181 LTLIDNRRSRAHPCSPMCKSGRCWSESSEDCOSLTRTCAGAGCARCKPLPDDCCEQC 240

QY 241 AAGCTGPKASDCLACLFHNSGICEIHCALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
 DB 241 AAGCTGPKASDCLACLFHNSGICEIHCALVTYNTDTFESMPNPEGRTYFGASCVTACP 300

QY 301 YNVLSTDVSGCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLGMOYIKANSKRTGIT 360
 DB 301 YNVLSTDVSGCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360

QY 361 ELSEFAGCKKIFGSLAFLEPESFDGDPASNTAPQEPQLQVFETLEETIGLYISAMPDLP 420
 DB 361 IOEFAGCKKIFGSLAFLEPESFDGDPASNTAPQEPQLQVFETLEETIGLYISAMPDLP 420

QY 421 DLSVQNLQVITGRILHNGAYSLLTQGLISWLGRLSRELGSGLALHHNTHLCFVHTV 480
 DB 421 DLSVQNLQVITGRILHNGAYSLLTQGLISWLGRLSRELGSGLALHHNTHLCFVHTV 480

QY 481 PMDQLFRNPHQALHTANRPEDECGEGGLAQCLARGCWGRPTQCNCSQFIRGQCC 540
 DB 481 PMDQLFRNPHQALHTANRPEDECGEGGLAQCLARGCWGRPTQCNCSQFIRGQCC 540

QY 541 VEECRVLOGLPREYVNAHRLCPHECQFQNSVTCFGEADQCVACAHYKDPFCVARG 600
 DB 541 VEECRVLOGLPREYVNAHRLCPHECQFQNSVTCFGEADQCVACAHYKDPFCVARG 600

QY 601 PEGVYKPDLSYMIKMFPEEGACQCPINCHSCVDLDKGCAPQORASPLTISAVVG 660
 DB 601 PEGVYKPDLSYMIKMFPEEGACQCPINCHSCVDLDKGCAPQORASPLTISAVVG 660

QY 661 SFMLRVPKYSASHLEKRRQOKIRKXTMRLLQETELVEPLTPSGAMPQAOQRILKETEL 720
 DB 661 ILLVVLGVPIGLIKRRQOKIRKXTMRLLQETELVEPLTPSGAMPQAOQRILKETEL 720

QY 721 RKVKVLGSGAFSTVYKGIWIPDGENVKIPLAIKVLRENTSPYANKELIDEAVYMGVSGP 780
 DB 721 RKVKVLGSGAFSTVYKGIWIPDGENVKIPLAIKVLRENTSPYANKELIDEAVYMGVSGP 780

QY 781 YVSRLLGICLSTVOLVQMPYGLLDHVENRRLSODLNMCMQIACMSGLTEVR 840
 DB 781 YVSRLLGICLSTVOLVQMPYGLLDHVENRRLSODLNMCMQIACMSGLTEVR 840

QY 841 LVHRDLAARNVLYKSPNVKITTDFGLARLLIDETEVYADGKVYIKMMALESIIRRF 900
 DB 841 LVHRDLAARNVLYKSPNVKITTDFGLARLLIDETEVYADGKVYIKMMALESIIRRF 900

QY 901 HOSDWSYGVYWEMLTGAAPYDGIIPAREIPLDLKEKERLPPOPICTIDVYMTWKVM 960
 DB 901 HOSDWSYGVYWEMLTGAAPYDGIIPAREIPLDLKEKERLPPOPICTIDVYMTWKVM 960

QY 961 IDSECRPRFREIVSEFSMARDPQRFVIVQNEIDGAPASPLDSTFRSLLEDMDGDLVDA 1020
 DB 961 IDSECRPRFREIVSEFSMARDPQRFVIVQNEIDGAPASPLDSTFRSLLEDMDGDLVDA 1020

QY 1021 EBYLVQGGFCPPDPAFGAGMVRHRRSSSTRSGGGLTGLPSEBEARSPPLAPSEG 1080

Query Match 96.9%; Score 6623; DB 23; Length 1255;

Db 1021 EEVLVPOGFFCPDDPAPGAGVHRRHSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 QY 1081 AGSDVFPDGLGMAKAGLQSTPTHDPSLQKYSDDPVLPSFTDGVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFPDGLGMAKAGLQSTPTHDPSLQKYSDDPVLPSFTDGVAPLTCSPQPEYV 1140
 QY 1141 NQPVPRPQPSRSGPLTPAARPAATLEBAKTLSPGNKGVKDVFAFGAVENPEYLTPO 1200
 Db 1141 NQPVPRPQPSRSGPLTPAARPAATLEBAKTLSPGNKGVKDVFAFGAVENPEYLTPO 1200
 QY 1201 GGAAAPQHPHPAFSPFNLVYVQDPPRCAPSTKGTPTAENPEYLTPO 1255
 Db 1201 GGAAAPQHPHPAFSPFNLVYVQDPPRCAPSTKGTPTAENPEYLTPO 1255

RESULT 13

AAM51143
 ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;

DT 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX tyrosine kinase; receptor; c-erbB2; gene therapy.

OS Homo sapiens.

XX Location/Qualifiers

FT Key 1..653

FT Domain /note= "extracellular domain"

FT Domain 676..1255

FT Domain /note= "intracellular domain"

FT Domain 990..1255

FT Domain /note= "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001MO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMK-) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI, 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX or enhancing an immune response to the protein, has Her-2/neu

XX extracellular domain fused to Her-2/neu intracellular or

XX phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

CC comprising the fusion proteins or nucleic acid molecules. In
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu
 CC protein is fused to a Her-2/neu intracellular domain or
 CC phosphorylation domain (or its Deltapart fragment). An immune
 CC response to Her-2/neu protein is elicited or enhanced by
 CC administering the fusion protein in the form of a vaccine, or by
 CC transfecting cells of an animal *ex vivo* with a nucleic acid
 CC encoding the fusion protein, and delivering the transfected cells
 CC to the animal. The fusion proteins, nucleic acids, and isolated
 CC specific T-cells are useful for inhibiting the development of a
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer
 CC in a patient. T cells that specifically react with a Her-2/neu
 CC fusion protein can be used to remove tumour cells from a sample in
 CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 23; Length 1255;

Best Local Similarity 97.3%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELALCRMGILALILPGAASTQVCTGDMKRLPASPETHLDMRLHYQGQVQGNL 60
 Db 1 MELALCRMGILALILPGAASTQVCTGDMKRLPASPETHLDMRLHYQGQVQGNL 60
 QY 61 ETTYLPNASTSLFQDIQEVQGVYLIANQVQVPLQRLIRYRGTLQFEDNYALAVLDNG 120
 Db 61 ETTYLPNASTSLFQDIQEVQGVYLIANQVQVPLQRLIRYRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVPGASPGGLRELQSLREILKGGVLIQNRNOLCYQDITLWMDIFHKNOLA 180
 Db 121 DPLNNTPTVPGASPGGLRELQSLREILKGGVLIQNRNOLCYQDITLWMDIFHKNOLA 180
 QY 121 DPLNNTPTVPGASPGGLRELQSLREILKGGVLIQNRNOLCYQDITLWMDIFHKNOLA 180
 Db 121 DPLNNTPTVPGASPGGLRELQSLREILKGGVLIQNRNOLCYQDITLWMDIFHKNOLA 180
 QY 181 LTLIDNRSRACHPCSPMKSGRCWSESESDCSLTRYCAGGCARCKPLPTDCHEQC 240
 Db 181 LTLIDNRSRACHPCSPMKSGRCWSESESDCSLTRYCAGGCARCKPLPTDCHEQC 240
 QY 241 AAGCTGPKSDCLACIHFHNSGICEJHCAVLTNTNTDTPESMNPFGRTTFGASCYTACP 300
 Db 241 AAGCTGPKSDCLACIHFHNSGICEJHCAVLTNTNTDTPESMNPFGRTTFGASCYTACP 300
 QY 301 YNLTSDVGSCTLVCELNQVETAEEDTQRCCKSPCARVCGGLGMEHREAVATSAN 360
 Db 301 YNLTSDVGSCTLVCELNQVETAEEDTQRCCKSPCARVCGGLGMEHREAVATSAN 360
 QY 361 ELEFACCKIIFGSLAFPSFDGDPASNTAPLOPQOVETLEETITGYLYISAWPDSL 420
 Db 361 ELEFACCKIIFGSLAFPSFDGDPASNTAPLOPQOVETLEETITGYLYISAWPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGSLREISGLLIHNTTLCEVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGSLREISGLLIHNTTLCEVHTV 480
 QY 481 PMDQFRNPQALHTANRPEDECEVGEGLACHOLCARGCHWGPPTQCVNCSQFLRGQEC 540
 Db 481 PMDQFRNPQALHTANRPEDECEVGEGLACHOLCARGCHWGPPTQCVNCSQFLRGQEC 540
 QY 541 VEECRVYQGLPREYVNAARHCLPCHPCQPNQSVTTFGBADQCVCAHAKBPFCVAC 600
 Db 541 VEECRVYQGLPREYVNAARHCLPCHPCQPNQSVTTFGBADQCVCAHAKBPFCVAC 600
 QY 601 PSQVFPDLSTMPIMKPPDEGACQPPINCTHSCVDLDKGCAPAEORASPLTIFNNFTV 660
 Db 601 PSQVFPDLSTMPIMKPPDEGACQPPINCTHSCVDLDKGCAPAEORASPLTIFNNFTV 660
 QY 661 SFULRVKVSASLHERRQOKIRKYMRRLLQETELVEPLTPSGAMPNOQMRILKETEL 720
 Db 661 SFULRVKVSASLHERRQOKIRKYMRRLLQETELVEPLTPSGAMPNOQMRILKETEL 720
 QY 721 RKRVKVGSGFGVYVGIWPDGENKIPAIIVLENTSPKXKXITLLEAYVAGVSGP 780
 Db 721 RKRVKVGSGFGVYVGIWPDGENKIPAIIVLENTSPKXKXITLLEAYVAGVSGP 780

QY 781 YVSRLLGICLTSTVQLTQMPYGCILDHRENRGSLGSDLLNMCQIAKGSYLEDR 840
 Db 781 YVSRLLGICLTSTVQLTQMPYGCILDHRENRGSLGSDLLNMCQIAKGSYLEDR 840
 QY 841 LVHRDLAARVVLKSPNHVKITDFGLARLLDIDETEHADGKVPIMKMALESILRRFT 900
 Db 841 LVHRDLAARVVLKSPNHVKITDFGLARLLDIDETEHADGKVPIMKMALESILRRFT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKEGRLPPPICTIDVYMIYKCM 960
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIDPLEKEGRLPPPICTIDVYMIYKCM 960
 QY 961 IDSECRPRFRELSEFSRMAKDPORFVIONEDLGPASPLDSTFYSLIEDDMGLVDA 1020
 Db 961 IDSECRPRFRELSEFSRMAKDPORFVIONEDLGPASPLDSTFYSLIEDDMGLVDA 1020
 QY 1021 EBYLVPOQGFCDPAPAGAGVYHHRSSSTRSGGDLTLGLEPSEEAAPSPLAPSEG 1080
 Db 1021 EBYLVPOQGFCDPAPAGAGVYHHRSSSTRSGGDLTLGLEPSEEAAPSPLAPSEG 1080
 QY 1081 AGSDVDFGDLGMAAGKLSLPTHDSPLOQYSEDPVLPBSTDGYVAPLTCSPQPEYV 1140
 Db 1081 AGSDVDFGDLGMAAGKLSLPTHDSPLOQYSEDPVLPBSTDGYVAPLTCSPQPEYV 1140
 QY 1141 NQPDVAPQPPSPREGPLPARPAGATLEBAKTLSPGKNGVYKDVFAFGAVENPEYLTPQ 1200
 Db 1141 NQPDVAPQPPSPREGPLPARPAGATLEBAKTLSPGKNGVYKDVFAFGAVENPEYLTPQ 1200
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDOPPERGAPPTFFKGTPTAENPEYLGLDVPV 1255
 Db 1201 GGAAPQHPPPAPSPAFDNLVYWDOPPERGAPPTFFKGTPTAENPEYLGLDVPV 1255

RESULT 14

AAU77114
 ID AAU77114 standard; Protein, 1255 AA.

AAU77114;

05-JUN-2002 (first entry)

Human Her-2/neu polypeptide.

Human; Her-2/neu; cytosolic; haematological malignancy; CML;

acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;

chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;

Hodgkin's lymphoma; T cell therapy.

Homo sapiens.

WO200213847-A2.

21-FEB-2002.

13-AUG-2001; 2001WO-US25408.

14-AUG-2000; 2000US-0638280.

28-SEP-2000; 2000US-0675904.

(CORI-) CORIXA CORP.

Gaiger A, Cheever MA, Hand-Zimmermann S,

WPI; 2002-280741/32.

N-PSDB; ABK10730.

Inhibiting haematological malignancy development by administering

polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide

encoding the polypeptide, or antigen presenting cells expressing the

polypeptide

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX
 SQ Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 23; Length 1255;
 Best Local Similarity 97.3%; Freq. No. 0;
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELALCMGILLALPPGAASVQVCTGDKMLRLPAPETHLMDLRHLYGQCVVQNL 60
 Db 1 MELALCMGILLALPPGAASVQVCTGDKMLRLPAPETHLMDLRHLYGQCVVQNL 60
 QY 61 ELYTPNASSLFLDIDIEVQGYVLIANQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
 Db 61 ELYTPNASSLFLDIDIEVQGYVLIANQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
 QY 121 DEINNTTGTASPGGLREIQLRSLETILKGVLIQRNPOLCYOTILMKDIFHNQOLA 180
 Db 121 DEINNTTGTASPGGLREIQLRSLETILKGVLIQRNPOLCYOTILMKDIFHNQOLA 180
 QY 181 LTLIDTNRSRACHPSPKCKSRGCESEEDQSLTRTVCAAGCARCKGLPTDCCHQC 240
 Db 181 LTLIDTNRSRACHPSPKCKSRGCESEEDQSLTRTVCAAGCARCKGLPTDCCHQC 240
 QY 241 AAGCTGPRHSDCLAFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 Db 241 AAGCTGPRHSDCLAFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
 QY 301 YNYLSTDVSGCTVCPHNCETVADGQRECKSKPARVCYGLGMYIANSRFTIT 360
 Db 301 YNYLSTDVSGCTVCPHNCETVADGQRECKSKPARVCYGLGMYIANSRFTIT 360
 QY 361 EIEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEOLQVFETLEITGYLISAMPDLP 420
 Db 361 EIEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEOLQVFETLEITGYLISAMPDLP 420
 QY 421 DLSVFNQLQVIRGSLIHNGANSLTLOGIGISLGLRSIREGSLALHNTHLCFVHTV 480
 Db 421 DLSVFNQLQVIRGSLIHNGANSLTLOGIGISLGLRSIREGSLALHNTHLCFVHTV 480
 QY 481 PMDQFRNFQALHTANRPEDECVGEGLACHOLCARGHCMPGTQCVNSQFLRGQEC 540
 Db 481 PMDQFRNFQALHTANRPEDECVGEGLACHOLCARGHCMPGTQCVNSQFLRGQEC 540
 QY 541 VEECRVLQGLREYVNAHCLPCHPECOQNGSVTCFPEADQCVACAHYKDPFCVARC 600
 Db 541 VEECRVLQGLREYVNAHCLPCHPECOQNGSVTCFPEADQCVACAHYKDPFCVARC 600
 QY 601 PSQVAPDSTYPIKFPDEBGACQPCPINTCHSCVDLDDKCPAEQASPLTISIAYVG 660
 Db 601 PSQVAPDSTYPIKFPDEBGACQPCPINTCHSCVDLDDKCPAEQASPLTISIAYVG 660
 QY 661 SFWLAVPVKASHLERKROOKIRKYTRRLQETELVEPLTPSGAMPNOQMRILKTEL 720
 Db 661 SFWLAVPVKASHLERKROOKIRKYTRRLQETELVEPLTPSGAMPNOQMRILKTEL 720
 QY 721 RKYVVLGSAFGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAAYVAGVSP 780
 Db 721 RKYVVLGSAFGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAAYVAGVSP 780
 QY 781 YVSRLLGICLTSTVQLTQMPYGCILDHRENRGSLGSDLLNMCQIAKGSYLEDR 840
 Db 781 YVSRLLGICLTSTVQLTQMPYGCILDHRENRGSLGSDLLNMCQIAKGSYLEDR 840
 QY 841 LVHRDLAARVVLKSPNHVKITDFGLARLLDIDETEHADGKVPIMKMALESILRRFT 900

```

Db      ||| 841 LVHRDLAARVNLVSPNHVKTDFGLARLLDIDETEHADGKVPFKMALESILRRFT 900
Qy      ||| 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEGERLPPQPICTIDVYMIWKCM 960
Db      ||| 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEGERLPPQPICTIDVYMIWKCM 960
Qy      ||| 961 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDMDGLVDA 1020
Db      ||| 961 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDMDGLVDA 1020
Qy      ||| 1021 EBYLVPOQGFPCPDPAAGAGMVAHHRSSSTRSGGDLTLGLEPSEEEAPRSPLAPSEG 1080
Db      ||| 1021 EBYLVPOQGFPCPDPAAGAGMVAHHRSSSTRSGGDLTLGLEPSEEEAPRSPLAPSEG 1080
Qy      ||| 1081 AGSDVFPDGLGMGAKGLQSLPTHDPSFLQRYSEDPVPLPSETDGVYVAPLTCSPOPEYV 1140
Db      ||| 1081 AGSDVFPDGLGMGAKGLQSLPTHDPSFLQRYSEDPVPLPSETDGVYVAPLTCSPOPEYV 1140
Qy      ||| 1141 NQDVRPQPSPREGSLPAARPAATLERAKTSPGNKGVYKDVAFGAVENPEYLTPO 1200
Db      ||| 1141 NQDVRPQPSPREGSLPAARPAATLERAKTSPGNKGVYKDVAFGAVENPEYLTPO 1200
Qy      ||| 1201 GGAAQPPHPPAPSPAFDNLVYMDQDPREGAPSTKGTPTAENPEYLGIDV 1255
Db      ||| 1201 GGAAQPPHPPAPSPAFDNLVYMDQDPREGAPSTKGTPTAENPEYLGIDV 1255

RESULT 15
ID AAR39568 standard; Protein: 1433 AA.
AC AAR39568;
DT 07-FEB-1994 (first entry)
DE Sequence of c-erbB-2 tumour antigen.
KM Tumour antigen; c-erbB-2; glycoprotein.
OS Homo sapiens.
PN W0316185-A.
PD 19-AUG-1993.
PF 05-FEB-1993; 93WO-US01055.
PR 06-FEB-1992; 92US-0831967.
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
PI Houston TX, Huston JS, Oppermann H, Ring DB,
XX WPI: 1993-272889/34.
XX N-PSDB; AAQ46083.
XX New single chain Fv polypeptide binding to C-erbB-2 tumour
XX antigen - for imaging or treating breast or ovarian cancer etc.
XX Disclosure; pages 48-54; 87pp; English.
XX c-erbB-2 refers to a protein antigen expressed on the surface of
XX tumour cells, such as breast and ovarian tumour cells, which is an
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
XX the location of a stop codon in AAQ46083.
SQ Sequence 1433 AA;
Query Match 96.3%; Score 6580; DB 14; Length 1433;
Best Local Similarity 96.7%; Pred. No. 0;

```

```

Matches 1214; Conservative 8; Mismatches 33; Indels 0; Gaps 0;
Qy      ||| 1 MELALCRKGLIALLPQAASTGYCTGTDMLRLPAPRETHLMDLRHLYOGCCQVYQNL 60
Db      ||| 1 MELALCRKGLIALLPQAASTGYCTGTDMLRLPAPRETHLMDLRHLYOGCCQVYQNL 60
Qy      ||| 61 ELYVLPNASLSFLDIOEVQGYVLIANQVQVPLQRLRIYRGTLPEFNALAVLNDG 120
Db      ||| 61 ELYVLPNASLSFLDIOEVQGYVLIANQVQVPLQRLRIYRGTLPEFNALAVLNDG 120
Qy      ||| 121 DELNNTPTVAGSPGRLRELQRLSTELLKGVLIQRNPOLCYQDTILMKDIFHNQOLA 180
Db      ||| 121 DELNNTPTVAGSPGRLRELQRLSTELLKGVLIQRNPOLCYQDTILMKDIFHNQOLA 180
Qy      ||| 181 LTLIDTNSRACHPCSPCKSGRCWSESEDCQSLTRVYAGGACRCKPLPTDCHEQC 240
Db      ||| 181 LTLIDTNSRACHPCSPCKSGRCWSESEDCQSLTRVYAGGACRCKPLPTDCHEQC 240
Qy      ||| 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPNEGRTTGASCVTACP 300
Db      ||| 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMNPNEGRTTGASCVTACP 300
Qy      ||| 301 VNYLSTDVSGCTLVGCPLNQEVTAEDGTQRCCKSPCARVCYGLMOYIKANSKFIGIT 360
Db      ||| 301 VNYLSTDVSGCTLVGCPLNQEVTAEDGTQRCCKSPCARVCYGLMOYIKANSKFIGIT 360
Qy      ||| 361 ELEFAGCKKIFGSLAFLESPFDGDPASNTAPLQBPOLQVFTLEBITGYLYISAMPDILP 420
Db      ||| 361 ELEFAGCKKIFGSLAFLESPFDGDPASNTAPLQBPOLQVFTLEBITGYLYISAMPDILP 420
Qy      ||| 421 DLSVQNLQVIRGLIHNGAYSLLTQGLISWLGRLSRLRELSGLALIHNNHLSFVHTV 480
Db      ||| 421 DLSVQNLQVIRGLIHNGAYSLLTQGLISWLGRLSRLRELSGLALIHNNHLSFVHTV 480
Qy      ||| 481 PMDQLFRNPHQALHTANRPEDECGEGIAHQCLCARGCMGPCTQCVNCSQFLRGQEC 540
Db      ||| 481 PMDQLFRNPHQALHTANRPEDECGEGIAHQCLCARGCMGPCTQCVNCSQFLRGQEC 540
Qy      ||| 541 VEECRVLOGLPREYVNAHCLPCHPECCOPNGSVTCFGBEADQVACAHYKDPFCVARG 600
Db      ||| 541 VEECRVLOGLPREYVNAHCLPCHPECCOPNGSVTCFGBEADQVACAHYKDPFCVARG 600
Qy      ||| 601 PSQVPPDLSYMPITWKFPEDEGACQPCPINCTHSCVDLDKGCAPQASPLTSTISAVVG 660
Db      ||| 601 PSQVPPDLSYMPITWKFPEDEGACQPCPINCTHSCVDLDKGCAPQASPLTSTISAVVG 660
Qy      ||| 661 SFMLRVKYSASHLEKRRQOKIRKYMRLQETELVEPLTPSGAMPQAOQRILKETEL 720
Db      ||| 661 SFMLRVKYSASHLEKRRQOKIRKYMRLQETELVEPLTPSGAMPQAOQRILKETEL 720
Qy      ||| 721 RYKXKLGSGAEGTYVKKGIWIPDGENVKIPVAIKYLRNTSPYANKEILDEAVVMGVGSP 780
Db      ||| 721 RYKXKLGSGAEGTYVKKGIWIPDGENVKIPVAIKYLRNTSPYANKEILDEAVVMGVGSP 780
Qy      ||| 781 YVSRLLGICLTSTVOLVTLQMLPYGCLLHVRENRGRLSGQDLNMCQIAKMSYLEDR 840
Db      ||| 781 YVSRLLGICLTSTVOLVTLQMLPYGCLLHVRENRGRLSGQDLNMCQIAKMSYLEDR 840
Qy      ||| 841 LVHRDLAARVNLVKSPPNHVKTDFGLARLLDIDETEHADGKVPFKMALESILRRFT 900
Db      ||| 841 LVHRDLAARVNLVKSPPNHVKTDFGLARLLDIDETEHADGKVPFKMALESILRRFT 900
Qy      ||| 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEGERLPPQPICTIDVYMIWKCM 960
Db      ||| 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEGERLPPQPICTIDVYMIWKCM 960
Qy      ||| 961 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDMDGLVDA 1020
Db      ||| 961 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDMDGLVDA 1020
Qy      ||| 1021 EBYLVPOQGFPCPDPAAGAGMVAHHRSSSTRSGGDLTLGLEPSEEEAPRSPLAPSEG 1080
Db      ||| 1021 EBYLVPOQGFPCPDPAAGAGMVAHHRSSSTRSGGDLTLGLEPSEEEAPRSPLAPSEG 1080

```

Qy	1081	AGSDVFDGDLGMAKGIQSLPTHDPSPLOKRYSEDPVPLJSEETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAKGIQSLPTHDPSPLOKRYSEDPVPLJSEETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVFRQPPSPREGPLPAARPAAGATLERAKTLPKXGKGVKDVFAFGAVENPEYLLTPQ	1200
Db	1141	NQPDVFRQPPSPREGPLPAARPAAGATLERAKTLPKXGKGVKDVFAFGAVENPEYLLTPQ	1200
Qy	1201	GGAAPOPHPPPAFSPAFDNLVYWDODPPREGAPPSTFKGPTAENPEYLGIDVPV	1255
Db	1201	GGAAPOPHPPPAFSPAFDNLVYWDODPPREGAPPSTFKGPTAENPEYLGIDVPV	1255

Search completed: July 22, 2003, 09:16:28
Job time : 41.9339 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.25 Seconds

(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-653-675-14

Perfect score: 6814
Sequence: 1 MELALCRMLLLALPPCA.....TFKGTPTANPEYLGLDVPV 1253

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6604	96.9	1255	1 A24571	protein-tyrosine k
2	5811	85.3	1260	1 TVRTNU	protein-tyrosine k
3	5801.5	85.1	1254	2 I48161	p-185 precursor -
4	3659	45.0	1210	1 GQHU	epidermal growth f
5	3040	44.6	1210	2 A53183	epidermal growth f
6	3017.5	44.3	1223	1 TVGHV	epidermal growth f
7	2908.5	42.7	1308	2 A47253	epidermal growth f
8	2598	38.1	1166	1 S6142	protein-tyrosine k
9	2380.5	34.9	1342	2 A36223	kinase-related tra
10	2292.5	33.6	1339	2 JC4387	epidermal growth f
11	1688.5	24.8	698	1 TVFVU	protein-tyrosine k
12	1626	23.9	604	1 TVYU	protein-tyrosine k
13	1599.5	23.5	1330	1 GQFE	epidermal growth f
14	1570	22.0	544	2 S3745	protein-tyrosine k
15	1563	22.9	545	2 S00727	kinase-related tra
16	1546	22.7	540	2 B44776	protein-tyrosine k
17	1544	22.7	540	1 TVFVB	protein-tyrosine k
18	1509	22.1	644	2 A36325	epidermal growth f
19	1487	18.9	1323	2 E88257	protein-let-23 (im
20	1487	18.9	1374	2 S70712	protein-tyrosine k
21	1194	17.5	1369	2 S70713	protein-tyrosine k
22	1142	16.8	1717	1 A45558	epidermal growth f
23	1126	16.5	527	2 A42032	epidermal growth f
24	977.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	724	10.6	1363	2 T43220	insulin-like growt
28	706	10.4	1383	2 A36090	insulin receptor p
29	705.5	10.4	1372	2 A34157	insulin receptor p

30	705	10.3	1382	1 INHUR	insulin receptor p
31	691.5	10.1	1300	2 A36502	insulin receptor-r
32	685	10.1	1477	2 T18534	protein-tyrosine k
33	680	10.0	1407	2 T43212	insulin-like growt
34	667	9.8	1268	2 B36502	insulin receptor-r
35	638	9.4	1367	1 IGHUR1	insulin-like growt
36	631.5	9.3	2148	1 A56081	insulin receptor -
37	625	9.2	2101	2 S57245	insulin receptor (
38	624	9.2	1371	2 A33837	insulin-like growt
39	622	9.1	1390	2 T10346	insulin receptor -
40	604	8.9	1114	1 S05582	protein-tyrosine k
41	592	8.7	987	2 A54092	protein-tyrosine k
42	584.5	8.6	952	2 I50612	protein-tyrosine k
43	584.5	8.6	1091	2 S33596	protein-tyrosine k
44	576.5	8.5	984	2 A39753	protein-tyrosine k
45	572.5	8.4	801	4 TVHURE	transforming prote

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: C-erb-B-2 protein precursor; Kinase-related transforming proteir

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #ext change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Senda, K.; Nomura, N.; Miyajima, N.; Saito, Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal grc

A/Reference number: A24571; MUID:86118653; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:K03163; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Senda, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T. Proc. Natl. Acad. Sci. U.S.A. 82, 6487-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/et

A/Reference number: A25491; MUID:86016729; PMID:2959567

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 717-1031 <SEM>

A/Cross-references: GB:M1167; NID:G182163; PIDN:AAA5809.1; PID:G553282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeb Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares c

A/Reference number: A44188; MUID:86070181; PMID:2939974

A/Accession: A44188

A/Molecule type: mRNA

A/Residues: 1-517; 'PALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A/Cross-references: GB:M11730; NID:G183986

R/King, C.R.; Kraus, M.H.; Aaronson, S.A. Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma

A/Reference number: I59509; MUID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 332-909 <REX>

A/Cross-references: GB:U29395; NID:G459807; PIDN:AAA5809.1; PID:G459808

R/Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D. Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HBR2 (neu) promoter: evidence for multiple mechanisms for transcripti

A/Reference number: I57622; MUID:87268898; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAL>

F:558-680/Domain: transmembrane #status predicted <TM>
 F:723-988/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:11,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:591/Binding site: phosphate (Thr) (covalent) #status predicted
 F:758/Active site: Lys #status predicted
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.3%; Score 5811; DB 1; Length 1260;
 Best Local Similarity 85.2%; Pred. No. 5, 1e-232;
 Matches 1074; Conservative 57; Mismatches 118; Indels 12; Gaps 4;

QY 1 MELALCRWGLLALLPFGAASVQCTGTDMLRLPASPEHLDMLRLHYGCGVQVQNL 60
 DB 4 MELAACRMWGLLALLPFGAASVQCTGTDMLRLPASPEHLDMLRLHYGCGVQVQNL 63
 QY 61 ELTYLPTNASLFLQDIOEVGYLILAHNOVROVPLQRLIRVGTQLEFEDNALAVLDNG 120
 DB 64 ELTYLPTNASLFLQDIOEVGYLILAHNOVROVPLQRLIRVGTQLEFEDNALAVLDNR 123
 QY 121 DPLNNTTPTVY-GASPGGLRELQRLSTELIKGGVLIQRPOLCYODTILMKDI FHKNNQL 179
 DB 124 DPGDNVAASPRGTPGRLRELQRLSTELIKGGVLIQRPOLCYODMVLKQVFRKNNQL 183
 QY 180 ALLILDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGRLPTDCHEQ 239
 DB 184 APVDIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGRLPTDCHEQ 243
 QY 240 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTAC 299
 DB 244 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMHNDEGRYTFGASCVTTC 303
 QY 300 PNYVLTSTVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCGYIGMOYIRANSKEFIGI 359
 DB 304 PNYVLTSTVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCGYIGMOYIRANSKEFIGI 363
 QY 360 TELEFAGCKKIFGSLFLPESFDGPDASNTAPLQPOLOVFEETLEITGLYISAMPDL 419
 DB 364 NVQEFQCKKIFGSLFLPESFDGPDASNTAPLQPOLOVFEETLEITGLYISAMPDL 423
 QY 420 PDLVFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRELGSGLALIHNTLHLCFVHT 479
 DB 424 RDLVFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRELGSGLALIHNTLHLCFVHT 483
 QY 480 VPDQVFRNPHQALLTANRPEDE-CVGBGLACHQICAGHCHGSPRTQCVNCSQTLRQ 538
 DB 484 VPDQVFRNPHQALLTANRPEDE-CVGBGLACHQICAGHCHGSPRTQCVNCSQTLRQ 543
 QY 539 ECVBECRVQLGLPREYVNAHCLPCHPECCOPNGSVTCGPEADQCVAAHYKDPFCVA 598
 DB 544 ECVBECRVWKGRLPREYVNAHCLPCHPECCOPNGSVTCGPEADQCVAAHYKDPFCVA 603
 QY 599 RCSSGVKPDLSYPIKPKPEDEACQPCPINCCHSCVDLDDKCCPAEQASPLTISVAV 658
 DB 604 RCSSGVKPDLSYPIKPKPEDEACQPCPINCCHSCVDLDDKCCPAEQASPLTISVAV 663
 QY 659 VGILLVVLGVFGIL-----FNNFTVSFMLRVKVASHSLEPTLPSCAMPNQOMKI 712
 DB 664 VGILLVVLGVFGIL-----FNNFTVSFMLRVKVASHSLEPTLPSCAMPNQOMKI 719
 QY 713 LKTELKRVKVLCSGAFGYKGIWIPDGENVKIPAIKYLRENTSPKANKETLDEAYYM 772
 DB 720 LKTELKRVKVLCSGAFGYKGIWIPDGENVKIPAIKYLRENTSPKANKETLDEAYYM 779
 QY 773 AGVSPVSVSLIGICLTSTVQVLTQMLPVGCLLDHRENRGRGLSGDGLNMWQIAKXMS 832
 DB 780 AGVSPVSVSLIGICLTSTVQVLTQMLPVGCLLDHRENRGRGLSGDGLNMWQIAKXMS 839
 QY 833 YLEDAVRLVHBDLAARVNLVKSNNHVKITDFGLARLLIDETEVHADGKGVPIKMALES 892
 DB 840 YLEDAVRLVHBDLAARVNLVKSNNHVKITDFGLARLLIDETEVHADGKGVPIKMALES 899
 QY 893 LRERFTHQSDVMSVGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 952

DB 900 LRERFTHQSDVMSVGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 959
 QY 953 MYKCMIMDSERPRRELVSFSSMARPPQRFVYVQNDLGPASPLDSTFYRSLLEDDDM 1012
 DB 960 MYKCMIMDSERPRRELVSFSSMARPPQRFVYVQNDLGPASPLDSTFYRSLLEDDDM 1019
 QY 1013 GDLYDAEYLVLPQCGFFCPDPAAGGVHHRHSSSTRSGCDLTLGLESEEEPRSP 1072
 DB 1020 GDLYDAEYLVLPQCGFFCPDPAAGGVHHRHSSSTRSGCDLTLGLESEEEPRSP 1079
 QY 1073 LAPSEGASDVFDGDLGMAKGLQSLPTDHPSELQRYSEDPYPLPSETGYAPLTC 1132
 DB 1080 LAPSEGASDVFDGDLGMAKGLQSLPTDHPSELQRYSEDPYPLPSETGYAPLTC 1139
 QY 1133 PQPEYVNOQVROPSPREGPLPAARAGATLBRAKTSLSGKGVVDFVAFGAVENP 1192
 DB 1140 PQPEYVNOQVROPSPREGPLPAARAGATLBRAKTSLSGKGVVDFVAFGAVENP 1199
 QY 1193 EYLTPQGAAPQHPHPPAFSPDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVP 1252
 DB 1200 EYLTPREGTASPPHSPAFSPDNLVYWDQNSSEQGPSPNFEGTPTAENPEYLGLDVP 1259
 QY 1253 V 1253
 DB 1260 V 1260

RESULT 3
 148161
 P-185 precursor - golden hamster
 C/Spectes: Mesocricetus auratus (golden hamster)
 C/Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C/Accession: 148161
 R/Nakamura, T.; Uehijima, T.; Ishizaka, Y.; Nago, M.; Arai, M.; Yamazaki, Y.; Ishi
 Gene 140, 251-255, 1994
 A/Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A/Reference number: 148161; MUID:94193007; PMID:7908275
 A/Accession: 148161
 A/Status: Preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-1254 <RES>
 A/Cross-references: GB:D16295; NID:9493236; PID:BAA03801.1; PID:9747595
 A/Genes: neu
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.1%; Score 5801.5; DB 2; Length 1254;
 Best Local Similarity 84.8%; Pred. No. 1, 2e-231;
 Matches 1068; Conservative 66; Mismatches 114; Indels 11; Gaps 3;

QY 1 MELALCRWGLLALLPFGAASVQCTGTDMLRLPASPEHLDMLRLHYGCGVQVQNL 60
 DB 1 MELAACRMWGLLALLPFGAASVQCTGTDMLRLPASPEHLDMLRLHYGCGVQVQNL 60
 QY 61 ELTYLPTNASLFLQDIOEVGYLILAHNOVROVPLQRLIRVGTQLEFEDNALAVLDNG 120
 DB 64 ELTYLPTNASLFLQDIOEVGYLILAHNOVROVPLQRLIRVGTQLEFEDNALAVLDNR 120
 QY 121 DPLNNTTPTVY-GASPGGLRELQRLSTELIKGGVLIQRPOLCYODTILMKDI FHKNNQL 180
 DB 124 DPLNNTTPTVY-GASPGGLRELQRLSTELIKGGVLIQRPOLCYODTILMKDI FHKNNQL 180
 QY 180 ALLILDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGRLPTDCHEQ 240
 DB 184 APVDIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGRLPTDCHEQ 240
 QY 240 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTAC 299
 DB 244 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTTC 300
 QY 300 PNYVLTSTVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCGYIGMOYIRANSKEFIGI 359
 DB 304 PNYVLTSTVGSCTLVCPRLHNOEYTAEDGTQRCCKSPCARVCGYIGMOYIRANSKEFIGI 363
 QY 360 TELEFAGCKKIFGSLFLPESFDGPDASNTAPLQPOLOVFEETLEITGLYISAMPDL 419
 DB 364 NVQEFQCKKIFGSLFLPESFDGPDASNTAPLQPOLOVFEETLEITGLYISAMPDL 423
 QY 420 PDLVFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRELGSGLALIHNTLHLCFVHT 479
 DB 424 RDLVFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRELGSGLALIHNTLHLCFVHT 483
 QY 480 VPDQVFRNPHQALLTANRPEDE-CVGBGLACHQICAGHCHGSPRTQCVNCSQTLRQ 538
 DB 484 VPDQVFRNPHQALLTANRPEDE-CVGBGLACHQICAGHCHGSPRTQCVNCSQTLRQ 543
 QY 539 ECVBECRVQLGLPREYVNAHCLPCHPECCOPNGSVTCGPEADQCVAAHYKDPFCVA 598
 DB 544 ECVBECRVWKGRLPREYVNAHCLPCHPECCOPNGSVTCGPEADQCVAAHYKDPFCVA 603
 QY 599 RCSSGVKPDLSYPIKPKPEDEACQPCPINCCHSCVDLDDKCCPAEQASPLTISVAV 658
 DB 604 RCSSGVKPDLSYPIKPKPEDEACQPCPINCCHSCVDLDDKCCPAEQASPLTISVAV 663
 QY 659 VGILLVVLGVFGIL-----FNNFTVSFMLRVKVASHSLEPTLPSCAMPNQOMKI 712
 DB 664 VGILLVVLGVFGIL-----FNNFTVSFMLRVKVASHSLEPTLPSCAMPNQOMKI 719
 QY 713 LKTELKRVKVLCSGAFGYKGIWIPDGENVKIPAIKYLRENTSPKANKETLDEAYYM 772
 DB 720 LKTELKRVKVLCSGAFGYKGIWIPDGENVKIPAIKYLRENTSPKANKETLDEAYYM 779
 QY 773 AGVSPVSVSLIGICLTSTVQVLTQMLPVGCLLDHRENRGRGLSGDGLNMWQIAKXMS 832
 DB 780 AGVSPVSVSLIGICLTSTVQVLTQMLPVGCLLDHRENRGRGLSGDGLNMWQIAKXMS 839
 QY 833 YLEDAVRLVHBDLAARVNLVKSNNHVKITDFGLARLLIDETEVHADGKGVPIKMALES 892
 DB 840 YLEDAVRLVHBDLAARVNLVKSNNHVKITDFGLARLLIDETEVHADGKGVPIKMALES 899
 QY 893 LRERFTHQSDVMSVGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 952

Nature 309, 270-273, 1984
 A>Title: ATP-stimulated interaction between epidermal growth factor receptor and superco
 A'Reference number: A38023; MUID:84191554; PMID:6325948
 A'Contents: annotation; receptor activity
 A'Note: The EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R'Chem: W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A>Title: Functional independence of the epidermal growth factor receptor from a domain
 A'Reference number: A3331; MUID:90003333; PMID:2790960
 A'Contents: annotation; internalization signal
 C'Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
 C'Genetics:
 A'Gene: GDB:EGFR
 A'Cross-references: GDB:120610; OMIM:131550
 A'Map position: 7p12.3-7p12.1
 C'Superfamily: epidermal growth factor receptor; protein kinase homology
 C'Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F1-24/Domain: signal sequence #status predicted <SIG>
 F1-25-1210/Product: EGF receptor #status predicted <EXT>
 F1-25-645/Domain: extracellular #status predicted <EXT>
 F1-75-100/Domain: EGF receptor extracellular domain repeat <EE1>
 F1-390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F1-646-668/Domain: transmembrane #status predicted <TM>
 F1-669-1210/Domain: intracellular #status predicted <INT>
 F1-10-975/Domain: protein kinase homology <KIN>
 F1-718-726/Region: protein kinase ATP-binding motif
 F1-999-1046/Region: coated-pit mediated internalization signal
 F1-1047-1210/Region: inhibitory
 F1-128-175,355,413,444,528,603/Binding site: carbohydrate (asn) (covalent) #status predic
 F1-745/Active site: Lys #status experimental

Query Match 45.0%; Score 3069; DB 1; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 3e-119;
 Matches 621; Conservative 175; Mismatches 361; Indels 108; Gaps 24;

```

QY 11 LLLALLPAGA--STVCCTGTMKRLPASPETHLDMHLKYGCCVQVQGNELTYLPTN 68
DB 14 LLAALCPARALEBKVKVCCGTGSKNLQGLTFEDHPLSLQRMNCEVULGNLEITVQGN 73
QY 69 ASLSFLQDIQEVQGVYLAHNOYRQVPLQRLIVRGTOLEFENYALAVINDGPDYNNTP 128
DB 74 YLSFLKTIQEVAGVYLAHNTVERPLENLOIRGNMYENSVALAVLSND----- 126
QY 129 VVGASPGRLREIQLRETELKGVILQGNPOLCYGDTLMDIFKNNQLALLIDTR 188
DB 127 ---ANKTGLKELPMRWLQELHGAVRFSNNPALCNVESTQMRDIVSSDLSNMSDPQNH 183
QY 189 SRACHPSPCKSGSRGCGSSSDQSLRTVCAAGCA-RCKGRLPTCCHEGCAAGCTGP 247
DB 184 LQSCQKCDSPCRGSCWAGBENCQKLTIKICAQCGSGRCGRKSPDCCHNCAAGCTGP 243
QY 248 KASDCLACLFPHNSGICELCPALVTYNTDFESMNPBGRYTFGASCTYACPYVYLTSD 307
DB 244 RSDCLVCKRFRDEATCKOTCPPLMYNTTYQMDVNPGRKSPGATCYKCKPRYVYVTD 303
QY 308 VGSCTLVCEPLHNOVETADGTORCEKSPCARVCGYLMQYIKANSKEITGTELE-FAG 366
DB 304 HSCVYACGADSEW-EDQVKKCKCKCEBPCRKVCNGIGIGERK-DLSLNNNTNKHFN 361
QY 367 CKKIFGSLAFPSFSGDPAANTAPQRPQLOVFTLEITGLYLSAAMPDLPDLSTVQ 426
DB 362 CTISGDLHLLPFAFGDSFTHTPPDLPQELDLIKVKEITGFLIQAMPERRDLDAHE 421
QY 427 NLOVIRGRIHNGAVSLTQGLGISWLGSRSLRELSGALIHNNHLCFVATVPDOLF 486
DB 422 NLEITIRGRKHQGFSLAVSLNITSLGRSLKEISDGVIIISGNLCLYANTIMKKLF 481
QY 487 RNPQALLHTANPEDECVGEGGLAQCLCARGHGCPGFTQCVNCSQITRGQECYEECKV 546
DB 482 GTSQGTXTKTIISNGENSCATGCVCHALCSPEGCWBPEDDVCVSCNVRGREGCVDKCL 541
QY 547 LOGLPREYVNAHCLCHPECPQNGSVTCFGEADQCAACAHYKDPFCVVRCPGSGVAP 606
  
```

```

DB 542 LEGEREFVENSECIQHPCECLPQAMNITCTGRGPDNQCIAHYIDGPHCKTCPCAGWG 601
QY 607 DLSMPLMKPDEEGACQPCPINCSTHSCVDLDDKGCAPBASPSTISVAVNG---ILL 663
DB 602 ENNTLVWKTADAGVCHLCHPNTCTCTGPGLEGGCTTGNPKP--SIATGMVALLLL 658
QY 664 VVVLGVVFGILLFNNFTV-SFWLRVPKYSASHLEPLTPSGAMPQACMRILKETLRKV 722
DB 659 VVALGI--GLFMRRIHVRKRLRLQLRELEVPPLTPSGEAPQALLRLIKETEFKKIK 716
QY 723 VLGGAGCTYKGIWIDGENVKIPVATKLRNTSKAKELLDEYLVAVAGVSPVYSR 782
DB 717 VLGGAGCTYKGIWIDGENVKIPVATKLRNTSKAKELLDEYLVAVAGVSPVYSR 776
QY 783 LLGICLSTVQVLTQMLPYGCLLDHVAENGRGLSQDYLWCMQIAKMSYLEDDVLR 842
DB 777 LLGICLSTVQVLTQMLPYGCLLDHVAENGRGLSQDYLWCMQIAKMSYLEDDVLR 836
QY 843 DLARNVLVSPNHVKTIDGRLALDIDETEHYADGKVPKMMALESLIRRFTHQSD 902
DB 837 DLARNVLVKTPOHVKTIDGRLALDIDETEHYADGKVPKMMALESLIRRYTHQSD 896
QY 903 VMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDYVMIKCMMDSE 962
DB 897 VMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDYVMIKCMMDAD 956
QY 963 CRPFREIVSEFSPBARDPQRFVYIQ-NEDLGPASPLDSTFFRSLEDDMDGLVDAEY 1021
DB 957 SRPFREILTEFSPBARDPQRFVYIQDERHGLSPFDSNRYALMBEDDDVDVDAEY 1016
QY 1022 LVPOQGFPCPDPAAGAGMHHNRSSSTSGGDDLTLGLEPSEEARPSLASREGGS 1081
DB 1017 LIPQGF-----SSPSTSRPPLSLSTST 1042
QY 1082 DVFDGDLQMGAKLQSLPTHDPSPRLQYSEDPVPLPSEF--DGVAFLPSCPOPEYVN 1139
DB 1043 N-NSTVACIDRNLOSGPIHDESFELGRYSDFGALTESIDTFL-----VPEVIN 1094
QY 1140 QPDVPRPPSPREGPLPAAPRAGATLEBAKLSGCKGKVQDVAFGGAENPEYL-TPQ 1198
DB 1095 Q-SVPRKPSAGVQHPVYHNOPLN-----APSRPHYOD--PHSTAVNPEYLVNQ 1143
QY 1139 GGAAPQHPAPFAPFAPNLVYMDQ-----DP-----PERGAPSTFGTPTAE 1242
DB 1144 -----PFCVNSTPDSPAHMAQKSHQSLNDPDYQDPFFKAKPKNGIFKGS-TAE 1193
QY 1243 NPEYL 1247
DB 1194 NAEYL 1198
  
```

RESULT 5
 A33183
 epidermal growth factor receptor precursor - mouse
 C'Species: Mus musculus (house mouse)
 C'Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C'Accession: A53183; A43818; S24942; A28941; S43325; 149643
 R'Luettiker, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N
 Genes Dev. 8, 399-413, 1994
 A>Title: The mouse waved-2 phenotype results from a point mutation in the EGF recep
 A'Reference number: A53183; MUID:94170986; PMID:815255
 A'Accession: A53183
 A'Molecule type: mRNA
 A'Results: 1-1210 <LUE>
 A'Cross-references: GB:U03425
 R'Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morze, B.
 Oncogene 6, 673-676, 1991
 A>Title: Comparison of EGF receptor sequences as a guide to study the ligand binding
 A'Reference number: A43818; MUID:91232866; PMID:2030916
 A'Accession: A43818
 A'Molecule type: mRNA
 A'Results: 1-714 <AVI>
 A'Cross-references: GB:X59598

R.Eisinger, D.P.; Serrero, G.
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S24942
 A:Accession: S24942
 A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:212608
 R:Heiserman, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; PMID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'K', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
 R:Heiserman, G.J.; Gill, G.N.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1971, 'X', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:g486830; PIDN:CAAS5587.1; PID:g486831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Day, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
 A:Reference number: 149643; PMID:93126380; PMID:7678348
 A:Accession: 149643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:g193001; PIDN:AAAS3029.1; PID:9567201
 C:Genetics:
 A:Gene: PGRF
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: Atp; growth factor receptor; kinase-related transforming protein; phosphop
 C:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <SIG>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase Atp-binding motif
 F:680-695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697-1070/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.6%; Score 3040; DB 2; Length 1210;
 Best Local Similarity 48.9%; Pred. No. 4, 7e-118;
 Matches 622; Conservative 168; Mismatches 369; Indels 114; Gaps 25;

364 -FACCKTIFGSLATPSPFDGDRASNTAPLOPGLQVETLEETIGLYISAMPDSDL 422
 Db HFKYCTAISGLHILPAFKGDSFTIRPPIDPRELEIKTKVETLGLLQAWDNDTL 417
 423 SVFQNLQVIRILHNGAVSLTLQGLISWLGRLSRELSGLALIHNNHLCFVHTVPW 482
 Db HAFENLEIRRTQHOFSLAVVGLNITSLGRLSLKEISGGDVITSGNRNLCAVNTINW 477
 483 DQLFRRPHQALLTARPEDECVSEGLAQHLCARGLCKWGRPGPQCNCSPFLRGSCVE 542
 Db KKLFGTPQKTKIMNNBAEKCKAVNVHNCPLCSSEGCWPEPEPDCVSCQNVSGRCVE 537
 543 ECRVLQGLPREYVNAHCLPCHRECPQNSVTSFCGPEADOCVACAYKDPFPCVACPS 602
 Db KQNLLEGRPREFVNSQCQHREPLCQANNITGTGGRPDNCICAHYIDGPHCVKCPA 597
 603 GVKPDLSTYPIWKFPEDEGACQPCFNGTSHCVLDLCKGPAEGRASPLTISVAVVGL 662
 Db GIMBENNTL-VMKYADANNVCHLCHANCYGCACPGLOQGEVWPSGPKIPSIATGIVGL 656
 663 LVVLGVVFGILF---NNEFVSWFLVVPKVSASHLEPLTPSGAMPQAOQRILKETELR 719
 Db LPIV-VALGIGLHRRRHIVKKTLLRLLOEBELVEPLTPSGAPNQHILRIKETEFK 715
 720 KVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKELIDEAVYMGVSPY 779
 Db KIKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKELIDEAVYMGVSPY 775
 780 VSRLLGICLTSTVQVLTQMLPYGCLLDHVRNKRGLSODLLNMCQIAKMSYLEDR 839
 Db VCRLLGICLTSTVQVLTQMLPYGCLLDHVRNKRGLSODLLNMCQIAKMSYLEDR 835
 840 VARDLAARNVLVKSPNHVKTDFGLARLDIDETEVADGGKVPKIMMALESILRRRFT 899
 Db VHRDLAARNVLVKSPNHVKTDFGLARLDIDETEVADGGKVPKIMMALESILRRRFT 895
 900 QSDVSYGVTVWELMTFCAGKYDGIIPAREIPDLLEKGRLEPQPICTIDVYIMVKKMT 959
 Db QSDVSYGVTVWELMTFCAGKYDGIIPAREIPDLLEKGRLEPQPICTIDVYIMVKKMT 955
 960 DSECPREELVSESRMARPOPFVVIQ-NEDGPAASPLSTFYRSLLEDMDGDVLDA 1018
 Db DSDSPKRELILEFSQVARDPQXIVYQSEBRNHLSPSTSNRYRALMDEDEDDVDA 1015
 1019 EBYLVPOQGFPCPPAPGAGMVHRRSSSTRSGGDLTGLPSEEBAPRSLAPSEG 1078
 Db DEYLVPOQGFPCPPAPGAGMVHRRSSSTRSGGDLTGLPSEEBAPRSLAPSEG 1074
 1079 AGSDVFDDDLGMGAAGLQSLPTHDPSPLOYSDDPTVPLPSET--DGTVAPLPCSPQPE 1136
 Db ATSN---NSTVACINRNGSCRKVEDAPLQYSSDPGAVEDNIDAF-----PVPE 1091
 1137 YVNPDPVAPPPSPREGPLPAAPAGATLEBAKTLSPQKGVVXDVAFAGAVENPEYL 1195
 Db YVNO-SVYKRPAGSVQNVYHNPQLHP-----APGRDHYQN--PHSNAVGNPEYL 1140
 1196 TPQGAQAQPHPPAPFSPAFDNLYVWD-----DP-----PERGAPPSFTKPTP 1239
 Db TAQ-----PTGLSSGFNSPALWIKGSHQMSLDNPDYQDFFPKETKNGIFKG-P 1190
 1240 TAENPEYGLDVP 1252
 Db TAENPEYGLDVP 1203

RESULT 6
 TVCHIV
 epidermal growth factor receptor precursor - chicken
 N:contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
 C:Species: Gallus gallus (chicken)
 C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
 C:Accession: A21720; A00643

R:Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart
Mol. Cell. Biol. 8, 1970-1978, 1988
A>Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Citterenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2989784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
C:Specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:31-307/Domain: EGF receptor extracellular domain repeat <EEL>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:336-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:92,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.3%; Score 3017.5; DB 1; Length 1223;
Best Local Similarity 47.9%; Pred. No. 4e-117;
Matches 621; Conservative 171; Mismatches 358; Indels 147; Gaps 27;

QY 8 RWGLALLALPPGAA-----STQVCTGTDMKRLPASPETHLMDLHLYQCCVQVGNLE 61
DB 13 RGAVALVLLLLGVALGSAVEKKYCOGTNNKLTQLGHVEDHFTSLQRMVNCCEVLNLE 72
QY 62 LTLPTNASTSFODIOEVGYVLAHNOVROYLQRLRYRGTOCFEDNYALAYLDNG 121
DB 73 ITVEHNRDLFTKLTIOEVAGYVLAHMDVLPLEVLQIRGNVLPDSFALAVSNTH 132
QY 122 PLNNTPTVTASBQGLRELQRLSTELKGVLIORNPOLCYODTILMDIFHKNQAL 181
DB 133 -MKKTQ-----GRELPMKRLSEILNGVAKISNPKLCNMDVTLAMDIDTSRK-PL 182
QY 182 TLID-TNRSACHPCSMCKGSRCKWSSSDCOSLTRTVAGGCA-RCKGRLPTDCHEO 239
DB 183 TVIDFASNLSSCPKCHNCTEDHCKWAGBQNCQTLTKVITCAQCCSGCKRKYVSDCHQ 242
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDPEPMNPEGRYTFGASCYTAC 299
DB 243 CAAGCTGPRSDCLACKKFRDADCTCOTCPVLVNTTYQMVNVEGKKSFGATCVREC 302
QY 300 PNYVLTSDVGSCTLVCLHNOETAEHGTORCEKSPCARVCYGLQMYTKANSFTGI 359
DB 303 PNYVVTVDHGSVCVRSCTDITYE-EENGVRKCKKCGGLCKVCNGIGIGLKILS-INA 360
QY 360 TEIE-FAAGCKIFGSLAFLEPSPGDPASNTAPLOEQLOVFELEITGYLISWMPDS 418
DB 361 TNDISFKNCTKINGDVSLPVALGDAFTKLPDLPRKLDVFTVEISGFLILQAMPDN 420
QY 419 LPDIASFQNTQVIRGRILHNGANSVLTQIGIGISMLGIRSLREISGLALIHNTLCPVH 478
DB 421 ATPLYAFENLEIRGRTRKQGYSLAVNLIKISLGRSKEXESDGDIALMKKNKLQYAD 480
QY 479 TVMDQLFRNPHQALLHTARPEDECVGEGLAGHQCARGHCGPPTQCVCNCSQFLRAG 538

DB 481 TMNRSPLFATOSQKTKIIONRKNKDCPADRHVCCPLCDVCGWGGPNCFCSCRFSSROK 540
QY 539 ECVBECRLTGLPREYVNAARHCLPCHRCOPONG---SVTFGPADCCVCAHYKDDPF 595
DB 541 ECVKQCNLLQBPREFERSKCLPCHSECLVQNSTANTTSSGPPDCMKCAHIDPH 600
QY 596 CVARCPSPKPDLSYMPKPEDEGACQCPINCTHSCVLDLDDKCPAQRASPLTSIV 655
DB 601 CVKAPPAVLBENDTL-VKXADANAVQQLCHPNCRTCKCPGLEGC---NGSKTFSIA 656
QY 656 SAVV-GILLVYLVGVFEILLFNNFTV-SFMLRPKVSASHLEPLTPSGAMPNQOMIL 713
DB 657 AGVGGALLCLVVGIGLILYRRRIYKRLRLLOSRELVETLPBSGAPNQAHRL 716
QY 714 KETELRYKVGSGAFGTVYKGIWIPDGENYKI PVAIKVLRNENTSPKANKELDEAYMA 773
DB 717 KETEFKRYKVLQSGAFGTVYKGIWIPBEXKIPVAIKELNENTSPKANKELDEAYMA 776
QY 774 GVSPEYVRLIGICLTSTVOLVTLQMPYGLLDHYRENRRGLSGODLLNCKQIAKMSY 833
DB 777 SYDNPHVGRLLGICLTSTVOLITQLMPPGCLLDYREKDMIGSGYLLNCKVQIAKMSY 836
QY 834 LEDVPLVNRDLAARVYLKSPNHVKITDPGLARLLDIDETEVHADGGKVPKIMALEST 893
DB 837 LEERLVRDLAARVYLKTPQHVKITDPGLAKLGADKSYHAGKVPKIMALEST 896
QY 894 RRRPTQSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIM 953
DB 897 HRIYHQSDWSYGVTVWELMTFGSKPYDGIPIAREISVLEKGERLPQPICTIDVYIM 956
QY 954 VKCMWIDECRRFELVSEFSRMAHDDQRFVIO-NDLGPASLDSTFRSLLEDDM 1012
DB 957 VKCMWIDDSRPFRELLAEFSKMAHDDQRFVIO-NDLGPASLDSTFRSLLEDDM 1016
QY 1013 GDLVPAEYIVPQGFPCPDPAAGAGVHHRHSSSTRSGGDTLLGLESEERAPSP 1072
DB 1017 EDIVYADEIVHQQFF-----NSPST-----SSTP 1042
QY 1073 L-----APSEGASVFPQDGLGAKAGLQSLPHDPSPLQRYSDPTVPLPSET--DGY 1125
DB 1043 LLSLSLSTSNNSATNCID-----RNGOGHPVEDPFPVORYSDPTGNFLESIDGIF 1094
QY 1126 VAPLFCSPQPEVNPDPVROPSPREBPLPAARAGATLEAKTSLSGKGVNDVF-- 1183
DB 1095 L-----PAPEYVNO-LMPKPS-----TAWQONIVNNISLT 1125
QY 1184 -----AEGAVENPEYLPQGAAPQHPAPPAPAFNDLYWDO----- 1223
DB 1126 AISKLPMDSRVYNSHSTAVDPEYL-----NTNQSPLAKTVFESSPYWIOSGNHOIN 1177
QY 1224 -DPE-----RGAPSTFKCTPTAENPEYGLDV 1252
DB 1178 LNDPDIQODFLNETKPNGLKVPAPENPEYLRVAP 1214

RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culnan, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A>Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PIO>
A:Cross-references: GB:L07868; NID:9337359; PIDN:AAB5946.1; PID:9337360
A>Note: sequence extracted from NCBI backbone (NCBI:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP, growth factor receptor
 F:716-981/Domain: protein kinase homology <Kin>
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 42.1% Score 2908.5; DB 2; Length 1308;
 Best Local Similarity 44.1%; Pred. No. 1,3e-112;
 Matches 600; Conservative 187; Mismatches 394; Indels 185; Gaps 28;

QY WGLLLALLPPGAA-----STQVCTGTDMKRLPASPEHLDMLRHLVYGGQVVOGNIETLY 64
 DB WWWVSLVLAAGTVQPSDSQSVCACTENKLSLSLDELQOYRALRKYEENECVVMGNLEITS 67
 QY LPPNALSIFLQDIOEOGYVLIHNOVROVPLORLIVRGTOLEFNRYALAVLDNDDELN 124
 DB IERNRDLSEFLRSREVTGYVVALNOFPYVLENLIIIRKTKYERVALATFLNKRKG 127
 QY 125 NTPVTVGASPDGLRELOLRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKONQALTLI 184
 DB 128 NF-----GLQELGLKNLTIELINGVYVDQNKELCYADTIHMODIVRNMPNSLTLV 178
 QY 185 DTRRSRACHPCSPCKSRGMSSEDCOSTRTVACAGC-AECCKPLPTDCCHCCAG 243
 DB 179 STNGSGGGRCHKSCG-RCMGPLENHQTLNRTVCABQDCGCGYVSDCCRRCAAG 237
 QY 244 CTGRKSDCLACLFHNSGICELHCPALVTYNTDTESMNPREGRYTFGASCVTAQPYN 303
 DB 238 CSGPKDTCFACMKNFNSGACVTCQCPQTFVYNPTTFLQENFNKTYGAFVCKKCPHNF 297
 QY 304 LSTDVSGCTVCPLEHNEVTAEDGTQCEKSKPCARVCYGLMQVITKANSKFTIGTELE 363
 DB 298 V-VQSSSCVACPSKMEV-EENGIRKCKKCTDIPCACGIGTGSMSQTVSSNIDK 355
 QY 364 FAGCKITFGSLAFIPESFGDPASNTAPLOPEQLQVETLEETITGYLYISAMPDLPJLS 423
 DB 356 FINTKINGNILEVITGHDPYALAIPEKLNVRITREITGFLINTOSWPNMTDE 415
 QY 424 VFQNLQVIRGRIHNGAYSLTQGLGISWLGSLRLSGGLAIHHNHTLCVHTVPM 483
 DB 416 VFSNLTVIGGRVLTYSGLSLILKQGITSLQFSLKEISAGNITYIDNSMLCYHTINMT 475
 QY 484 QLFENPHQALHTNRPEDCEVGEGLACHOLCAHGMGPGTQCVNCSQFLGOECVEE 543
 DB 476 TLFSTINORIYIRNRKAENCTAEGMCMNHCSSDCGMGPGPOCLSCRFSGRITIES 535
 QY 544 CRVLQGLPREVYNAHCLPCHPECQ-ONCSVTGFGSEADQVCAHAYKDPPECVACPS 602
 DB 536 CNLYDGEFREFENSGICVECDPQCEKMEGDLTCHGPGPNCTKSHFKGPGNVEKCPD 595
 QY 603 GVRPDLGYMPWKPEDEGACQPCPINCTHSCVLDLQKQ-----PAEQRASPL 651
 DB 596 GLQGANST--LFXADDPRECHPCHPCTQGCNGPTSHDCLYFTWGHSTLPQHR-TPL 652
 QY 652 TSISAVV-GILLVVVLGVFGLIFNNFTVFWLRVYKVSASH-----LEBLPSGA 703
 DB 653 --IAAGVIGLFIIVGLFPAVYVRK-----SIKKKALRFLTELVEPLTSPSGT 703
 QY 704 MPNOQVRILKETELRKKTLSGSAFGTVYKGIWIPGENTKIYALIKVLRNTPSPANK 763
 DB 704 APNOQVRILKETELRKKTLSGSAFGTVYKGIWIPGENTKIYALIKVLRNTPSPANK 763
 QY 764 EILDAAVYVAGVSPVYSLIGLICLTSTVQVLTQIMPYGCLLDHVRNRRGLSGODLLN 823
 DB 764 EFMDEALIMAGMDPHLVRLLGVCLSPITQVLTQIMPGCLLEVEHKKONIGSOLLN 823
 QY 824 CMQIAKMSIYEDVRLVHRDLAARNVLYKSPNHYKITDFGLARLLIDETRYHADGKVP 883
 DB 824 CVQIAKMSIYEDVRLVHRDLAARNVLYKSPNHYKITDFGLARLLIDETRYHADGKVP 883
 QY 884 IKMVALBSILRRRFTHSDVMSYGVYVWELMTFGAKPYDGIPIAREIDPLEKGERLQPP 943
 DB 884 IKMVALBSILRRRFTHSDVMSYGVYVWELMTFGAKPYDGIPIAREIDPLEKGERLQPP 943
 QY 944 ICTIDVYIMVCKMWIDSECRPREELVSEFSRKARDQRFVVIQNEB-LGPASPLDSTF 1002

DB 944 ICTIDVYIMVCKMWIDSECRPREELVSEFSRKARDQRFVVIQNEB-LGPASPLDSTF 1003
 QY 1003 YRSLEDDKMDLVDAEYLYVPOGFPDPAAGAMHRRSSSTRSGGLTIGLE 1062
 DB 1004 FQNLDEEDLEDDMDAEYLYV-QAFNIPPP-----IYSSRAIDNRS-----EIGHS 1051
 QY 1063 PSEBAPPS-----PLAP-SEGAGSDVFDGLGMGAA 1093
 DB 1052 PPAVTPMSQVFPYRDGFAEAGVSVYPAFTSTIPEAVVAGATGATIEIDSCNGL 1111
 QY 1094 KGLSLPTHPDSPLORYSEDEPTVLP-S-----ETDGYVAPLTCSPOPEYVNOBVRPQ 1146
 DB 1112 RKPVAHVQEDSSTQRYGADPTVFAPEKSPRGELDEEGYMTKMRDKPQOYLYNVE 1167
 QY 1147 PPSREGLPAPAPAGATLEBAKTLSPKKNVVDVFAFGAVENPEYLTPOGGAPOPH 1206
 DB 1168 -----ENPFVSR-----KNDLQ-----ALDNPEIHANSNG----- 1194
 QY 1207 PPPA-----FSPAFLNLYYMDQDPPERGA-- 1230
 DB 1195 PPXADEVYNEPLVNTFANTLGAKEYLKNILSMPEKAKKAFDNDYMHSLPFRSTLQ 1254
 QY 1231 PPSTFKGTP-----AENPEYL 1247
 DB 1255 HPDYQEVSTKYFYKQNGRIRPIVANEPEYL 1285

RESULT 8

S06142
 protein-tyrosine kinase (BC 2.7.1.112) mrk-y precursor - southern platyfish
 N:Alternate names: epidermal growth factor receptor homology; kinase-related transfo
 C:Species: Xiphophorus maculatus (southern platyfish)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
 R:Accession: S06142; S13809
 R:Hitbrodt: J; Adam, D; Malleschek, B; Maeueter, W; Raulf, F; Telling, A; Ro
 Nature 341, 415-421, 1989
 A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
 A:Reference number: S06142; MUID:90015140; PMID:2797166
 A:Accession: S06142
 A:Molecule type: DNA
 A:Residues: 1-1166 <MT>
 A:Cross-references: EMBL:X16891; NID:965220; PID:CAA34770.1; PID:9652291
 R:Adam, D; Maeueter, W; Scharf, M.
 Oncogene 6, 73-80, 1991
 A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
 A:Reference number: S13807; MUID:91125882; PMID:1846957
 A:Accession: S13809
 A:Status: preliminary; translation not shown
 A:Molecule type: DNA
 A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
 A:Cross-references: EMBL:X56319; NID:965284; PID:CAA39763.1; PID:965285
 C:Gene: mrk
 A:Gene: mrk
 A:Map position: Y
 A:Intons: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP, growth factor receptor; phosphotransferase; transmembrane protein;
 F:125/Domain: signal sequence #status predicted <Sig>
 F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MT>
 F:707-972/Domain: protein kinase homology <Kin>
 F:715-723/Region: protein kinase ATP-binding motif

Query Match 38.1% Score 2598; DB 1; Length 1166;
 Best Local Similarity 44.1%; Pred. No. 7e-100;
 Matches 561; Conservative 168; Mismatches 394; Indels 148; Gaps 31;

QY 4 AALCEWGLLALLPPGAAST-----QVCTGTDMKRLPASPEHLDMLRHLVYGGQVVOGNI 59
 DB 8 AALLQ--LLIVLSISRCSTDPDRKVCQSTSNQMT---LDNHYLKMKKMVGSCVAVLEN 62
 QY 60 LETTYPTNASLSFLQDIOEOGYVLIHNOVROVPLORLIVRGTOLEFNRYALAVLDND 119


```

Db      63 LEITYQENQDLSFLQSIQEVGVYLIANNEVSTIPLVNLRLIRGQNLTEGNTLLVMSN 122
Qy      120 GBLNNTTAVTASPCGLRELOLRSLTEILKGVLIQRNPCLCYODTILMKDIFHKNNL 179
Db      123 YOK-NBSSP--DYQVGLKQQLQSLNTEILSGVKSHPPLCNVETIIMWIDVDTSNP 179
Qy      180 ALTLIDTNSRACHPCSPMKCSRCWSESSEDCOSLTRTVCAAGC-ARCKGLPLPDCCHE 238
Db      180 TNNLIPHAFERQOQKCDHGCNVGSCWAPGRHCQKFTKLCABQCRRRCRGRKPIDCCNE 239
Qy      239 QCAAGCTGPKHSCLLCLFHNHSGICELHCPALVYNTDFFSMRPEBRATFFGASCYTA 298
Db      240 HCAAGCTGGRATDCLACRDNFDGCTKDCPEPKYDIVSHVYVNPNTKYFFGACVYE 299
Qy      299 CPNVLSTVGSCTTLCPLHNOEVAEDGTQCEKSCPKCARVCYGLQWYIKANSKFIG 358
Db      300 CPENYVTE-GACVRCSCSAGMLEVD-ENKRSCKPDCGCPVCCSIGLSL-SNTIAN 356
Qy      359 ITEL-EPAGCKTIFGSLAPLPSFDDPASPNTAPLOPELOVFEITLITGLYISAMPD 417
Db      357 STNIRSFNCTKINGDIIINRNSFEEDPHYKIGTMDPEHMLTLVKEITGLYVIMMPE 416
Qy      418 SLPDLSPVONLOYTRGRILHNGAYS-LTQGLISWLGRLRELSGLALHNTHLCE 476
Db      417 NMTLSVFNQLEIRGRTTFSRGSFVYVQVRLHQLGRLSKVAGNVILKNTQLRY 476
Qy      477 VHTVPDOLFRNHQALLHTANPEDECEVGLACHQLCARHGWPGPTQCVNCSQPLR 536
Db      477 ANIMRRLFRSEDSQIEVDART-----ENQTCNNESEDCGW-PCPTMVCSCILHVR 528
Qy      537 GQCEVECEVLOGLPREVYANARHCLPHECCPQNSVYCFGEADQCAVCAHYDPRPC 596
Db      529 GGRCVASCNLQOEPRBAOVDRGVCHQECVQDLSLTQCYGGRANCKSAHFQGPQC 588
Qy      597 VACPSGVNPDLSYMPIMKFPDEEGACQPCPINCTSHCVLDLDDKGPAPQASPLTIS 656
Db      589 IPRCPHGLIGDGGTL-IMKYADMGCCQCHQCTGCGGPGLSGCRGP-IVSHSLANG 646
Qy      657 AVAGILLVVLGVVFGIL-----INNFVYSFWLKPXKASHLBEUTSSGAMPQAKR 711
Db      647 LVSGLLITVALIIVLLRRRRIRKRRIRKRLLOKEL---VEBLTSSGAPQAFIR 702
Qy      712 ILKETELRKVYLGSGAFGVYKGIWIPDGENVKIPIVAIKVIRENSPRANEILDEAY 771
Db      703 ILKETELPKDRVLGSGAFGVYKGLNPPGENTRIPVAILKVLRENSPRVQGEVDEAY 762
Qy      772 MAGVSPFYVSLIGITLSTVQVLTQLMYGCILDHVRENRGLSGQDILLMCQIATKM 831
Db      763 MASVDHPHVCRLIGILTSAVOLVTQLMYGCILDYVRQHOERIQQWLLMNCVOIATKM 822
Qy      832 SYLEDVRLVHRDLAANVLYKSPNHVITDFGLARLDIDETEHYHADGGKVPILKMALES 891
Db      823 NYIEERHLYHRDLAANVLYKSPNHVITDFGLARLDIDETEHYHADGGKVPILKMALES 882
Qy      892 ILERRPTHOSDVMSYGVYWEIMTFSKAPYDGIIPAEIPDLLEKGRPLPQPICTIDVM 951
Db      883 ILOMTYTHOSDVMSYGVYWEIMTFSKAPYDGIIPAEIPDLLEKGRPLPQPICTIDVM 942
Qy      952 IMVKCMVIDSECPRELVSEFSRMAPDPPFVYVQNDLGPASLDTFNRSLIEDD 1011
Db      943 IILKCMVIDSSFRPRELVSEFSRMAPDPPFVYVQNDLGPASLDTFNRSLIEDD 999
Qy      1012 MGDVLAEEYLVQGFCEPDPAAGGVNHRHRSSTRSGGDLTLGLEBSEBAPRS 1071
Db      1000 --DVVDADAYLLPYKRI-----NRGS-----E 1020
Qy      1072 PLAPSEGASDVVDGLGMAKGLQSLPTHSPYQRYSEDPTV-PLPSEMDGYAPLT 1130
Db      1021 PCLPPTGTH-----PYRENSITLRNISPQNALKEKDDG----- 1055
Qy      1131 CSFQPEYVNPQVYRPP-----PSPRE-----GPLP--AARPAAGATLERAKTLLSPKNG 1177
Db      1056 -----EYVNPQGETSSRLSDIYNPYVEDLTQMGVSLSSQEAETNFSRPREYLMNTNONS 1110

```

```

Qy      1178 VKQVFAFGAVENPEVLTPOGGAAPQHPPEPAPSDNLYWDODPPEKAPPSTFKG 1237
Db      1111 L---PYVSSGSMDDPDY---QAG-----YQAAF-----LPQALGALUNGMF 1145
Qy      1238 TPTAENPEYLG 1248
Db      1146 LPAENLEAYLG 1156

RESULT 9
A:6223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; 159164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popsescu, N.C.; Axelson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERB3, a third member of the ERBB/epider
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <RES>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor recept
A:Reference number: 159164; MUID:90311312; PMID:2164210
A:Accession: 159164
A:Status: preliminary; translated from GB/EML/DDBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:G183990; PID:AAA35979.1; PID:G306841
A:Genetics:
A:Gene: GDB:ERB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <XIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match      34.9%; Score 2380.5; DB 2; Length 1342;
Beet Local Similarity 40.0%; Pred. No. 7.1e-91;
Matches 526; Conservative 195; Mismatches 454; Indels 141; Gaps 35;

Qy      10 GLTALTPPGA--STQVCTGTDMLRLPASPEHLDMLRLYQGCQVYQNLVLYLT 67
Db      11 GLTFLARGSEVNSQAVCPGLTNGLSVTGDAENQVLYLTYRCEVWGNLEIVLGH 70
Qy      68 NASLSFLDIDIEVQGYVLIANQVRYQVPLQRLTYRGQLFEDNYALAVLDNGPINTT 127
Db      71 NADSLFLOMIREVYGVYLVANNEFSTPLPPLRVLRGVQVYDGKFAIFVM-----LNNT 125
Qy      128 PVTGSPGGLRELOLRSLTEILKGVLIQRNPCLCYODTILMKDIFHKNNLALTLDITN 187
Db      126 ----NSSHALRQLALTQVTEILSGVYIEKDKXCHMDTIDMRIVADRO---AEIVKD 176
Qy      188 RSRACHPCSPMKCSRCWSESSEDCOSLTRTVCAAGC-ARCKGLPLPDCHEQCAAGCTG 246
Db      179 NGRSCPPCHVEVCKG-RCWGPSEEDCQTLTKTICAPQCNHCFGNPNQCCHEBCAGCSG 237
Qy      247 PKHSDCLACLHFNHSGICELHCPALVYNTDFFSMRPEBRATFFGASCYVACPYNLYST 306
Db      238 PQLDTCFACRHFNDSGACVPRCPQLVYNTKLTFLDENPHKTYQYGVGVASCPHNY-V 296
Qy      307 DVGSCCTVCPHNOEVAEDGTQCEKSCPKCARVCYGLQWYIKANSKP--IGITEL- 363
Db      297 DQTSVYACPPDKXLEV-DKNGLKWCCEPCGGLCPACACGTG-----SSSRFQTVSSNIDG 350
Qy      364 FAGCKTIFGSLAPLPSFDDPASPNTAPLOPELOVFEITLITGLYISAMPDLS 423

```

Db 351 EVNCTKILCNLDPLITGLNGDPWHKI PALDPEKLVNFTVREITGYLNI QSPRPHMHNS 410
 Qy 424 VFQNLQVIRGRILHNGAYS-LTLOGIGISWGLRSRLRELSGLALIHNTLHCFVHTVPM 482
 Db 411 VTSNLTIGRSLYNNGFSLIMKNLNTVSLGFRSLKESAGRIYISANRQLCYHNSLNM 470
 Qy 483 DQLEFRPHQALHTA-NRPEDECVGGLACHOLCARGHGWMGGPQCVCNCSQFLGQEGVC 541
 Db 471 TKYLRGPTEERLDIKNRRPRDCAVAGKVCPLCSSGGCGPQGLCSNYSRGGVCV 530
 Qy 542 EECRVQGLPREYVNAHCLPCHPECOPONGSVTCGPEPADOCVACAHYKDEPFCVACP 601
 Db 531 THCNFLNGERERFNAHECFSCHPCEQMEGTATNCSSSDICACAHRRDPRHVCSCP 590
 Qy 602 SGVCPDLSTYPIKPKPDEEGACOPPCINCHSCVDLDDKCAEQA---SPLSYSA 657
 Db 591 HGVLG-AKGPVYKPDVQNECPCHENCTQCGKPELDCCGLQTLVLGKTHLTMALTV 648
 Qy 658 VVGILLV-VVLGVVF---GILFNNFTVSFWLVRVKVASHLEPLTPSGAMPNQAKRI 712
 Db 649 IAGLVIFMMLGSTFLYWRGRRIQNKRAKRYE---RGESEPLDPS-EKANKVLARI 703
 Qy 713 LKETLRKVKVSGAGFYVYKGIWIPDGENVKIPVAIKYLRNTPFKANKELDEAYVM 772
 Db 704 FKETEIRKLVKLGSGVFGIVHKGMWIPBESIKIPVCIKVIEDKSGRQGFQAVTDMAL 763
 Qy 773 AGVSPVSRLLGICLTSTVQVLTOLMPYGLLDHRENRGRGSGODLNMGCQIKGMS 832
 Db 764 GSDIHAIVRLGICPSSSLQVLTQYIPLGSLLDHVRHGLGPPQLNMWQIKMGV 823
 Qy 833 YLEEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDDETEYHADGKVPKIMALESI 892
 Db 824 YLEHGMVHRNLAAARNVLLKSPQVQVADGVADLLPPDQKQLYSEAKTPKIMALESI 883
 Qy 893 LRRFTHQSVWVSGVTWELMTFGAPYDGIAPAREPDLLEKGEPLPPECTIDVWM 952
 Db 884 HFGKYTHQSDVWVSGVTWELMTFGAPYDGIAPAREPDLLEKGEPLPPECTIDVWM 943
 Qy 953 MVKCMWIDSECRPRELVEFSRMAPDQRFVVIQNEDELPA---SPIDSTFYRSLLED 1009
 Db 944 MVKCMWIDSECRPRELVEFSRMAPDQRFVVIQNEDELPA---SPIDSTFYRSLLED 1002
 Qy 1010 DMGMDLVDAEYLVPOGFCPPDPAFAGGVMHHRHSSSTRSGGGLTLGLEP-SEEA 1068
 Db 1003 VELEPELDLDLEAEDD---NLATTLTGSAISLPVGLTMRPG 1043
 Qy 1069 PRSPLASSEGASGVFDGLGMGAKGLQSLPTHD-PSPLQRYSEDPVTLF---SE 1121
 Db 1044 SQSLSPSSGV-WMNQNGNIGESQCSAVSGSSERCPVSLH-----EMPRGLASES 1096
 Qy 1122 TDGVYA-----PLTCSPOPE---YNQPDVRRQPSRREGP----- 1154
 Db 1097 SEGHVTSSEAELOEKVMCSRSRSRPRPGSAYHSQHSLLTPPTLSPGLEEDV 1156
 Qy 1155 ---LPARAPGATLEBAKTLSP-GKNGVY---KDVAFGAGVANNPEYLTPQGAAP 1203
 Db 1157 NGYMPDTHLKGTSSREGTSSVGLSVLSTEEDD-----EYEVNMRRRHNSP 1208
 Qy 1204 QHPPPAFSPAFLNLYWD-----QDPPERGAPPSSTKGIPTLANNPEYL 1247
 Db 1209 -PHPPRSSLEELGYEYMDVSGDSLASLGTSQCLHPVPVIMPTAGTPDEDEYVM 1263

RESULT 10
 J04387
 epidermal growth factor receptor homolog precursor - rat
 N/A: alternate names: ErbB3 protein; HER3 protein
 C: species: Rattus norvegicus (Norway rat)
 C: date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
 C: accession: J04387
 R: Heiliger, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
 Gene 165, 219-284, 1995
 A: title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A: reference number: J04387; MUID: 96096535; PMID: 8522190
 A: accession: J04387
 A: molecule type: mRNA
 A: residues: 1-1339 <HEL>
 A: cross-references: GI:U29339; NID:9915389; PID:9915390
 A: experimental source: liver
 A: note: The authors translated the codon AAC for residue 369 as Thr and GTT for res;
 C: comment: This protein is a functional heregulin receptor that transduces signals t
 C: gene: ErbB3
 C: superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
 C: keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote
 F: 1-19/DNA: signal sequence #status predicted <SIG>
 F: 20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
 F: 640-659/DNA: transmembrane #status predicted <TM>
 F: 705-970/DNA: protein kinase homology <KIN>
 F: 713-721/Region: protein kinase ATP-binding motif
 F: 939,1051,1156,1196,1257,1259,1273,1286,1325/Binding site: phosphate (TY
 Query Match 33.6%; Score 2292.5; DB 2; Length 1339;
 Best Local Similarity 40.2%; Pred. No. 2.9e-87;
 Matches 517; Conservative 171; Mismatches 434; Indels 165; Gaps 36;

Qy 3 LAALCRWGLLALLPRGA---STVCTGTMTKRLPASPTLDMRLHYOGGOVQGN 59
 Db 7 LQVLC---FLSLARSGEMNSQAVCGTNGSLVGDADNQYQTLKLYEKEGVWGN 62
 Qy 60 LELTFLPTNASTLFDIOEYQGYVLIHNOVROPVLPQRLRIYVGTOLFEDNVALAVLDN 119
 Db 63 LEIVLTGNADSLFQWIREVTAYVAVVAMNEFSLPRLNLAHVAGTQYDQKFAIFV-- 120
 Qy 120 GDPINNTPTVTAAPGSLREQLRSLEILKGVLIQRNPOLCYOCTILMKDITHKQNL 179
 Db 121 ---LNVNT---NSSHALRQLKFTQLTLEILSGVYIEKNDKLCHMDTIDWDIVRV-- 170
 Qy 180 ALTLIDTNRSAHCPSMCKGSRMGSSSDCOSLTPRYVAGGC-ARCKGRLPTDCHE 238
 Db 171 GAEIVYKNGNANCPCHVEYCKG-RCMGSGPDDQLITITICAPQCNCGCFEPNPGCHD 229
 Qy 239 QCAAGCTPKKSDCLACLFHNSGI:CEILHCPALVTYNTDTESMNPREGRTFGASCYTA 298
 Db 230 ECAGGCGSGHDDCFACRFNDGACVPRCEPLVYNNKLTQLEBNPHTKQYGGVAVAS 289
 Qy 299 CPYNNLTDSVGSCTLVCLHNGEYAEAGTCRCEKSPCARVYGL--GMQYIANKSF 356
 Db 290 CPNHFV-VQGFVCAKPPDKKEVD-KHGLMKCECGGLCPKACGCTSGSRYQVDSN 347
 Qy 357 IGITELFAGCKRIFGSLAFLPESFGDPASNTAPLOEOLQVETLEITGYLYISAP 416
 Db 348 ID---GFVNCCTKIIGNDPLITGLNDPWHKI PALDPEKLVNFTVREITGYLNIQSWP 403
 Qy 417 DSLPDLISVQNLQVIRGRILHNGAYS-LTLOGIGISWGLRSRLRELSGLALIHNTLH 475
 Db 404 PMHNFVSFNLITTGRELNYRGSLLIMKNLNTVSLGFRSLKESAGRIYISANRQL 463
 Qy 476 FVHTVPMDDLFNPHQALHTA-NRPEDECVGGLACHOLCARGHGWMGGPQCVCNCSQ 534
 Db 464 YHSLNMTKLLGSPBEERLDIKYRPLGCLAEKGVCPPLSGGCGMGAPOGLCSNRY 523
 Qy 535 LRQDECVESRYQLPREYVNAHCLPCHPECOPONGSVTCGPEPADOCVACAHYKPP 594
 Db 524 SREGVCTVHCNPLQSGPREFVHAQCFSCHPCELPMEGSTYNGSGSACACAFHFRGP 583
 Qy 595 FCVAPRGSGVXKPDLSMPYKPPDEGACOPPCINCHSC--VDLDDKCAEQAQASPLT 652
 Db 584 HCVNSCPHGLIG-AKGPVYKPDVQNECPCHENCTQCGKPELDCCGLQTLVLGKTH 641
 Qy 653 STVASAVGI-LVVLGVVF---GILFNNFTVSFWLVRVKVASHLEPLTPSGAMPNQ 707
 Db 642 LVLAATVGLAVIIMLIGSFLYWRGRRIQNKRAKRYE---RGESEPLDPS-EKANK 696
 Qy 708 AQMRILKETELRKVVLGSGAGFYVYKGIWIPDGENVKIPVAIKYLRNTPFKANKEL 767

F:130-395/Domain: protein kinase homology <kin>
 F:138-146/Region: protein kinase ATP-binding motif
 F:165/Active site: lys #status predicted

Query Match 23.9%; Score 1626; DB 1; Length 604;
 Best Local Similarity 50.7%; Pred. No. 3.5e-60;
 Matches 349; Conservative 74; Mismatches 140; Indels 126; Gaps 16;

QY 587 CAHYKDPFCVAKCPGKVPDLSYMPKFPDEGACQPPINCHSVCYDLDDKGCPEAQ 646
 DB 3 CAHFIDGPHCVKACPGAVLGENDTL-VKRYADANAACQLCHNCTRGCKGPGEGCP-- 58
 QY 647 RASPLTSYSAVY-GILLVVVIGVVGIIIFNNFTV-SPMLRVPKVSASHLEPLTPSGAM 704
 DB 59 NSKTSPIAGVGVGLCLVVGSLGIGLYRRHRYKRTLRLOERELVEPLTPSGA 118
 QY 705 PNOQMRILKETELRKVKVLSGAFSTYVKGWIPDGENVKIPVAIKVLRNTSPYANKE 764
 DB 119 PNOAHRIKETEFKKVKVLSGAFSTYVKGWIPDGENVKIPVAIKVLRNTSPYANKE 178
 QY 765 ILDEAYVMAGVSPYVSRILGTLSTVQVLTQMLMPYGLDHRNRRIGSQDILNMC 824
 DB 179 ILDEAYVMASVNDPHVCRILGTLSTVQVLTQMLMPYGLDHRNRRIGSQDILNMC 238
 QY 825 MOIAKMSYDEDELVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 884
 DB 239 VOIAKGMNVLERRLVHRDLAARNVLKTPQHVKITDFGLAKLLGADKEXYHAEKGVPI 298
 QY 885 KMALESILRRRTHQSDVMSYGVYVMTLTPGAKPRDGI-PAREIDLEKGRRLPQPI 944
 DB 299 KMALESILRRRTHQSDVMSYGVYVMTLTPGAKPRDGI-PAREIDLEKGRRLPQPI 358
 QY 945 CTIDVYIMVKCMWIDSECRPRELVEFSRMAPRPFVVIQ-NEDIGPASPLDSTFY 1003
 DB 359 CTIDVYIMVKCMWIDSECRPRELVEFSRMAPRPFVVIQ-NEDIGPASPLDSTFY 418
 QY 1004 RSLIEDDMDGLVDAAEYLYPQGGFCPPAPAGAGMHHRRSSSTRSGGDLTIGLEB 1063
 DB 419 RSLIEDDMDGLVDAAEYLYPQGGFCPPAPAGAGMHHRRSSSTRSGGDLTIGLEB 449
 QY 1064 SEERAPRSP-----ASEEGASDVPFDGLGMAKAGLQSLPHDPSPLQRYSEDPVPL 1118
 DB 450 -----SRPLLSLSATSNNNSATKCID-----RNGQHPREDSFVRYISDPTGNF 496
 QY 1119 PSET-DGYVAPLTCSQPEYVQPDVPPQPSFREGPLPAPAPAGATLERAKTUSPGKN 1176
 DB 497 LEESIDGFL-----PAEYVQ--LMPKKPSTAM----- 524
 QY 1177 GYVADVAF-----GGAVENPEYLTGQGAAPQPHPPAPAFADN 1217
 DB 525 -VONQITNFISLTAKLPMDSRYONSHSTAVDPEYL-----NTNOSPILAKTYFES 575
 QY 1218 LYVWDQDPPERGAPSTFKGTPTAENPEY 1246
 DB 576 SPYWIQSGNHQ-----INDNDY 594

RESULT 13
 GQFE
 epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
 N/Contains: protein-tyrosine kinase (BC 2.7.1.112) erpB
 C/Species: Drosophila melanogaster
 C/Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
 C/Accession: A00640; A38021
 R/Lytle, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B. Z.
 Cell 40, 559-607, 1985
 A/Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
 A/Reference number: A00640; MUID:85124611; PMID:2982499
 A/Accession: A00640
 A/Molecule type: DNA
 A/Residues: 1-1330 <liv>
 A/Cross-references: EMBL:K03054
 R/Masworth, S. C.; Vincent III, W. S.; Bilodeau-Wentworth, D.

Nature 314, 178-180, 1995

A/Title: A Drosophila genomic sequence with homology to human epidermal growth factor

A/Reference number: A38021; MUID:85137938; PMID:2983232

A/Accession: A38021

A/Molecule type: DNA

A/Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <MAD>

A/Cross-references: EMBL:X02293; NID:97922; PIDN:CAA26157.1; PID:9929565

C/Comment: This sequence is tentative because the introns have not been identified.

C/Genetics:

A/Genes: FlyBase:Bgfr

A/Cross-references: FlyBase:FBgn0003731

A/Map position: 2.57P

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph

F:1-732/Domain: extracellular #status predicted <EXT>

F:733-764/Domain: transmembrane #status predicted <TM>

F:765-1330/Domain: intracellular #status predicted <INT>

F:808-1072/Domain: protein kinase homology <kin>

F:816-824/Region: protein kinase ATP-binding motif

F:122-300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat

F:174/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic

F:181/Binding site: lys #status predicted

F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

Query Match 23.5%; Score 1599.5; DB 1; Length 1330;

Best Local Similarity 29.3%; Pred. No. 9.4e-59;

Matches 406; Conservative 177; Mismatches 415; Indels 389; Gaps 40;

QY 80 VQGYVLAMNQVRPRLRIRIVGTQF-----EDNALVALVDGPELNTTPTVGTASP 134
 DB 38 ITNYVIGLIDLPCTLSRLQIRRTLSVSEBKALV-----TY 81
 QY 135 GGLREQLRSLTEILKGVLIQRNPOLCYDTIIMKDI-FHKNQALATLIDTNSRACHP 194
 DB 82 SMYTLLEIPDLRDVLNGVGFHNNYNLCHMETIQSBIVSGTADYVYVDFAPARECP 141
 QY 195 CSPMKGRKWESEDDQSLRTYACAGCA--RKGPLPDDCHCECAAGTGRKSDC 252
 DB 142 CHESCTHG-CWGEGRKQCKRSKLTCSQCGRGYGRKPECCFLPAGGCTGTQKDC 200
 QY 253 IACLPFNSGICELCPALVTYNTDFESMPREGRYFGASCYVACEVNYLTDSVGSCT 312
 DB 201 IACKNFDEAVKECECPMKRKNPTTYLTENPEKVAYGATCVACEP-GHLLRNGACV 259
 QY 313 LVCPHNDVETADGTCRCCKSPCAVCGVGLGMQYIKANSKFGITEL-----EPAG 366
 DB 260 RSCPDQKMDKGE-----CVPNGCPCKPTC-----PGVTVLHAGNIDSEFN 300
 QY 367 CKKIFGSLAFESFDG--DPASNTA-----PLQEOUOVETLEITGYIYISAMPDS 418
 DB 301 CTVIDGNIRILDQTSRGPDVYANTMGPRYIFLDPERREVSIVKEITGYINIGTHPQ 360
 QY 419 LPDLVFNQLQVIRGRILHNGAY-SLTQGLGISWLGRLSEELGSLALHHNTHLCEV 477
 DB 361 FNLISFRLLETIRQLMESWFAALATVKSLSVLEMRNLKQISSGSVYIOHNRDLCY 420
 QY 478 HTVPMDQLFRNPHQALHTANREDEC----- 504
 DB 421 SNIRPAIQKEPEQKWNENLRADLCGKFTLLISVGNITIMHFAICREKMHLLSV 480
 QY 505 ----- 504
 DB 481 QGRLLGSHGSAVPLQELQFWHLHRLMLYQVINSIQKSNHQLTDACSPSVP 540
 QY 505 -----VG 506
 DB 541 SLTTERARYAIGAGLAMELEQITASASMRHSKTLPAEGROPRWVFLGVCASAPAGIA 600
 QY 507 EGLA-----CHQICARGHCWPGPTQCVNCSQFLRGCEVCEGRVQLGPREVY--N 556
 DB 601 EPLAGAVORKCHPLCELCTNYGYHGVQSKTTHYKREQCCTEC-----PADYTDSE 654
 QY 557 ARHCLPCHPECCQPNQSVTCFPEADQCVACHYK-----DPFF-----CVARCSG 603

```

Db      655 QRECFQRHPEC---NG---CTGGADCKSCSRNFKLFDANETGPRVYNSMTMFCSTKCPLE 708
Qy      604 VK-PDLSYPMFKPDEEGACQPCPINCCHSCVDLDCQCPAEQASPTLSVSAVGL 662
Db      709 MRHVNQYTAIGPY-----CAASPRSSKITANDL-----VNNFIITIGAV 749
Qy      663 LVVVLGVFGILF-----NNFTVSPMLRVPKVSASHLEPTLSPGAMPNOAKRILE 715
Db      750 LVPTTICLCVTTICGKOKAKKETVMMALSGRDS--EPLRPENIGANLCKLRIVVD 807
Qy      716 TELRKVKVLSGAFGTYYKGIWIPDENVKIPVAIKVLRNTSPKANKEILDEAYVAV 775
Db      808 AELRKGGVLMGAFGRTYKGVWVPEGENVKIPVAIKELIKSTGABSESEFLEAYIMASE 867
Qy      776 GSPYVRLLGICLTSTVOLVTCMLPFGCLLDHYENRGRLLGSODLNNMCQAKGMYLE 835
Db      868 EHVNLKTLAVCHSSQMLITQMLPGLCLLDYKNNRDKIGSKALLNMSTQJAKGMYLE 927
Qy      836 DVLVLRDLAARNVLYK---SPNHVKITDGLARLLDIDETEXHADGKVPKIMWALESI 892
Db      928 EKLVHRDLAARNVLYRLLAGEDH---DFGLAKLLSDSNEYKAKGKPIKMLALECI 983
Qy      893 LRRRFTHOSDVWYGVYTWELMTFGAKPYDGIPIAREIPDLLEKGEPLPQPICTIDYMI 952
Db      984 RNVFTSKSDVMAFGVITMELLTFGQRPHEHNPAKDIPDLIEVGLTEQPEICSLDIYCT 1043
Qy      953 MYVCMWIDSECRPFRELYSEFSRMARDPQRFVIONEDLG--PASPLDSTFYRSLLIED 1010
Db      1044 LLSCHMLDAMRPTFQGLTTFVFAEPARDGRIYALIGDKFTRIPA-----YTSQDEKD 1096
Qy      1011 ---DMGDLVDAEYLVPOQGFPCPDPAAGKGVHHRHSSSTRSGGDLTGLDPESEE 1067
Db      1097 LIRKLAPTTDSGLAIKKPDYLOPKALGPS-----HRTDCI-----DE 1135
Qy      1068 AP-----RSPLASGASGSDVPG--DLGMAKAGLQSLPTHDSPLQRYSEDPVPL 1116
Db      1136 MPTLNRYCKDPSKHSSTGDDERDSSAREVGVNLR-----IDL 1174
Qy      1119 PSTEDGYVAPLTCSPQPEYVNOVDVPPQPSPREGLPAPAPAGATLEAKTLPQKNGV 1178
Db      1175 PVDEDDYLP--TCQPPNNNNNNNN-----NPNQNNMAAVGVAAGY----- 1214
Qy      1179 VKGVFAFGAGVENEPEYL---TPQGAAPQPH-----PPAPSP-AF 1215
Db      1215 --DLIGVPSVDNPEYLLNAQTILGVGESPLPTQTIGLPMVGGCTMEVAVPMGSEPTSS 1272
Qy      1216 DNLXYWD 1222
Db      1273 DHEYYND 1279

```

RESULT 14

S35745
 protein-tyrosine kinase (EC 2.7.1.112) erbb - avian erythroblastosis virus
 C/Species: avian erythroblastosis virus
 C/Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C/Accession: S35745
 R/Vennstrom, B.
 submitted to the EMBL Data Library, March 1993
 A/Reference number: S35743
 A/Accession: S35745
 A/Molecule type: DNA
 A/Residues: 1-544 <VEN>
 A/Cross-references: EMBL:X12707
 C/Genetics:
 A/Gene: erbb
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 F/135-400/Domain: protein kinase homology <KIN>
 F/143-151/Region: protein kinase ATP-binding motif
 F/170/Active site: Lys #status predicted

```

Query Match      23.0%; Score 1570; DB 2; Length 544;
Best Local Similarity 53.3%; Pred. No. 6,46-58;
Matches 334; Conservative 68; Mismatches 133; Indels 92; Gaps 15;

Qy      578 GPEADQVACHAHYDPPFCVARGSPGVKPDLSYPMFKPDEEGACQPCPINCCHSCVDL 637
Db      1 GP--DHCKKCAHFDIGPCHVACAPAGVLENDTL--WKYADANAVCQCHNCTRGGCKGP 57
Qy      638 DQKGPARGASPLTSISAVV--GILLVTVGVVFGILIFNNFTV--SPMLRVPKVSASHL 695
Db      58 GLEGC---NSKTPSIAGVVGGLLVVGLGILYLRHRHVRKRTLRLLDQERLV 114
Qy      636 EPLTPSGAMPNOAKRIKETELRKVKVLSGAFGTYYKGIWIPDENVKIPVAIKVLR 755
Db      115 EPLTPSGAMPNOAHLRIKETEFKVKVLSGAFGTYYKGIWIPDENVKIPVAIKVLR 174
Qy      756 NTSPKANKEILDEAYVMAVGSPPVSRLLGICLTSTVOLVTCMLPFGCLLDHYENRRL 815
Db      175 ATSPKANKEILDEAYVMAVSDNPHVCRLLGICLTSTVOLVTCMLPFGCLLDHYRHKONI 234
Qy      816 GSODLNNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEX 875
Db      235 GSQVLYNNMCQIAKGMNVLERRHWHRLAARNVLYKTPQHVKITDGLAQGLADEKEY 294
Qy      876 HADGKVPKIMWALESIILRRRFTHOSDVWYGVYTWELMTFGAKPYDGIPIAREIPDLLEK 935
Db      295 HAEGKVPKIMWALESIILRRRFTHOSDVWYGVYTWELMTFGSKPYDGIPIAREISVLEK 354
Qy      936 GERLPQPICTIDYVIMVWKCMMIDSECRPFRELYSEFSRMARDPQRFVIONEDLG-- 994
Db      355 GERLPQPICTIDYVIMVWKCMMIDSECRPFRELYSEFSRMARDPQRFVIONEDLG-- 414
Qy      995 ASPLDSTFYRSLLIEDDMDGLVDAEYLVPOQGFPCPDPAAGKGVHHRHSSSTRSGG 1054
Db      415 PSPSTSKYRTIMEEDMEDIVDAEYLVPOQGF-----NSPST----- 454
Qy      1055 GDLTIGLEPSEERARSP-----APSGASDVFDGDLGMAKAGLQSLPTHDSPLQR 1109
Db      455 -----SRTLLSSLSKSTNSNTNCIDRNG-----H----- 481
Qy      1110 YSEDPVLPSETDGYVAPLTCSPQPEYVNOVDVPPQPSPREGLPAPAPAGAT-LEBA 1168
Db      482 -----PYREDFGL-----PAPEYVNO--LMPKKPSTAMVQNOYNYISLTAJSL 524
Qy      1169 KTLSPGKGVKQVAVFAFGAVNPEYL 1195
Db      525 PIDSRYN-----SHSTAVDNPEYL 544

```

RESULT 15

S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.1.-) - avian erythroblastosis vi
 C/Species: avian erythroblastosis virus
 C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C/Accession: S00727
 R/Scottling, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A/Title: Common site of mutation in the erbb gene of avian erythroblastosis virus m
 A/Reference number: S00727; MUID:88217326; PMID:2897102
 A/Accession: S00727
 A/Molecule type: DNA
 A/Residues: 1-545 <SCO>
 A/Cross-references: EMBL:X06943
 C/Genetics:
 A/Gene: erbb
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; phosphotransferase
 F/135-400/Domain: protein kinase homology <KIN>
 F/143-151/Region: protein kinase ATP-binding motif

Query Match 22.9%; Score 1563; DB 2; Length 545;
 Best Local Similarity 53.3%; Pred. No. 1,36-57;
 Matches 334; Conservative 67; Mismatches 134; Indels 92; Gaps 15;

```
QY 578 GPEADOCVACAHYKDPFVCVACPSGVKDDLSIMPIWKFPDEBACQPCPINCTHSCVDL 637
Db 1 GP--DHMKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAACOLCHPNCTRCCKGP 57
QY 638 DDGCEAECASPLTSIVSAVV-GILLVVVLGVVFGILLFNNFTV-SFWLAVPKVASHL 695
Db 58 GLEGCP--NGSKTPTSIAGVVGGLCLVVGIGIGLYLRRIHVKRTLRLQRELY 114
QY 696 EPLTPSGAMPNOQMRLKETELKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVIRE 755
Db 115 EPLTPSGEAPNOAHLRIKETEFKVKVLSGAFGVYKGLWIPESKVTIPVAIKELRE 174
QY 756 NTSBKANKETLDEAVYVAGVSPYVSRLLGICLSTVOLVTOLMPYGCLLDHYREGRLL 815
Db 175 ATSPKANKETLDEAVYVAGVSPYVSRLLGICLSTVOLVTOLMPYGCLLDHYREGRLL 234
QY 816 GSODLNMCMQIAKMSYLEDEVLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEY 875
Db 235 GSQYLLNMCMQIAKMSYLEDEVLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEY 294
QY 876 HAOGKVPITKMALESLRRRFTHOSVWGYVTWELMTFGAKPYDGI PAREIPDLLEK 935
Db 295 HAOGKVPITKMALESLRRRFTHOSVWGYVTWELMTFGAKPYDGI PAREIPDLLEK 354
QY 936 GERLPQPICTIDVYIMVCKMMIDSECRPRELVSEFSRMARDPQRFVVIQ-NEDLGP 994
Db 355 GERLPQPICTIDVYIMVCKMMIDSECRPRELVSEFSRMARDPQRFVVIQ-NEDLGP 414
QY 995 ASPLDSTFYRSLLEDDMGVLVDAEYLVPQCGFPCDPAPGAGMTHHRSSSTRSGG 1054
Db 415 PSPTDSKFKRTLMEEDMEDIVDADEVLPHQGF-----NSPST----- 454
QY 1055 GDLTLGLPSEEEAPRSP-----APSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLOK 1109
Db 455 -----SRTPLLSSLSLATSNNSATINCIDRNG-----H----- 481
QY 1110 YSEDPVPLPSETDGYVAPLTCSPQPEYVNOVDVRPQPSPREGPLFAAPAGAT-LERA 1168
Db 482 -----PVREDGFL-----PAPRYVNO-LMPKKPSTAMVONQIYVYISLTAISKL 524
QY 1169 KTISPKNQVVKDVFAGFAGAVENPEYL 1195
Db 525 PMDSRYQN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:27:42
Job time : 30.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.625 Seconds

(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-653-675-14

Perfect score: 6814
Sequence: 1 MEALALCRMGLLALLPAGA.....TKGTPPAENPEYLGLDVPV 1253

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6604	96.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5817	85.4	1257	1 ERB2_RAT	P06494 rattus norv
3	5801.5	85.1	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3067	45.0	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3041	44.6	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	2908.5	42.7	1308	1 ERB3_HUMAN	Q15303 homo sapien
7	2889	42.4	1308	1 ERB3_RAT	Q62356 rattus norv
8	2614.5	38.4	1157	1 XMRP_XIPMA	P13388 xiphioporus
9	2389.5	35.1	1132	1 ERB3_HUMAN	P21660 homo sapien
10	2315.5	34.0	1132	1 ERB3_RAT	Q62799 rattus norv
11	1924	28.2	1426	1 EGFR_DROME	P04412 drosophila
12	1672.5	24.5	634	1 ERBB_ALV	P00534 avian leuko
13	1626	23.9	604	1 ERBB_AVIER	P00535 avian eryth
14	1553	22.8	540	1 ERBB_AVIER	P11373 avian eryth
15	1518	22.3	703	1 EGFR_CHICK	P13387 gallus galli
16	1287	18.9	1323	1 LTR2_CABEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2_MOUSE	P70424 mus musculu
18	724	10.6	1363	1 ILPR_BRALA	Q02466 brachyosco
19	706	10.4	1363	1 INSR_RAT	P15127 rattus norv
20	705.5	10.4	1372	1 INSR_MOUSE	P15088 mus musculu
21	705	10.3	1300	1 INSR_MOUSE	Q9w114 mus musculu
22	703	10.3	1382	1 INSR_HUMAN	P06513 homo sapien
23	692.5	10.2	1380	1 INSR_HUMAN	P14616 homo sapien
24	685	10.1	1297	1 INSR_HUMAN	P14616 homo sapien
25	685	10.1	1477	1 HTR7_HYDAT	Q25197 hydat atten
26	680	10.0	1607	1 MIPR_LYMAT	Q25410 lyntaea stea
27	638	9.4	1367	1 IGR_HUMAN	P08069 homo sapien
28	625	9.2	1373	1 IGR_MOUSE	Q06071 mus musculu
29	625	9.2	2146	1 INSR_DROME	P09208 drosophila
30	622	9.1	1390	1 INSR_MOUSE	Q93105 aedes aegypt
31	621.5	9.1	1370	1 IGR_RAT	P24062 rattus norv
32	604	8.9	1114	1 RET_HUMAN	P07949 homo sapien
33	592	8.7	987	1 EPB4_HUMAN	P54760 homo sapien

34	584.5	8.6	984	1 EPB1_CHICK	Q07494 gallus galli
35	576.5	8.5	984	1 EPB1_RAT	P09759 rattus norv
36	573	8.4	902	1 EPB2_XENLA	Q01736 xenopus lae
37	572.5	8.4	977	1 EPB2_MOUSE	Q03145 mus musculu
38	572	8.4	987	1 EPB4_MOUSE	P54761 mus musculu
39	570.5	8.4	984	1 EPB1_HUMAN	P54762 mus musculu
40	569	8.4	1068	1 HTR6_HYDAT	P53356 hydat atten
41	569	8.4	1068	1 FAK1_XENLA	Q00944 gallus galli
42	568.5	8.3	1053	1 FAK1_CHICK	P34152 mus musculu
43	567	8.3	1052	1 FAK1_MOUSE	Q91571 xenopus lae
44	565.5	8.3	985	1 EPB1_XENLA	Q03346 rattus norv
45	562	8.2	1055	1 FAK1_RAT	

ALIGNMENTS

RESULT 1
ID ERB2_HUMAN STANDARD; PRT; 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p15erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor (HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86118663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234 (1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=2999974;
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139 (1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low U., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429 (1993).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTH: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M11767; AAA35808.1; JOINED.
 DR EMBL; M11761; AAA35808.1; JOINED.
 DR EMBL; M11762; AAA35808.1; JOINED.
 DR EMBL; M11763; AAA35808.1; JOINED.
 DR EMBL; M11764; AAA35808.1; JOINED.
 DR EMBL; M11765; AAA35808.1; JOINED.
 DR EMBL; M11766; AAA35808.1; JOINED.
 DR EMBL; M11730; AAA35493.1; JOINED.
 DR EMBL; M12036; AAA35978.1; JOINED.
 DR EMBL; X03363; CAA27060.1; JOINED.
 DR PIR; A25491; A25491.
 DR PIR; A24571; A24571.
 DR HSSP; P11362; IFGK.
 DR Genew; HGNC:3430; ERBB2.
 DR MIM; 164870; ERBB2.
 DR Interpro; IPR000494; EGFR_L_domain.
 DR Interpro; IPR000719; Euk_Dkinase.
 DR Interpro; IPR002174; Furin-like.
 DR Interpro; IPR001245; Tyr_Dkinase.
 DR Interpro; IPR004019; YIP_motif.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YIP; 2.
 DR ProDom; PD000001; Euk_Dkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TyKc; 1.
 DR ProSite; PS00107; PROTEIN_KINASE_ATP; 1.
 DR ProSite; PS00109; PROTEIN_KINASE_TYR; 1.
 DR ProSite; PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 720 987 PROTEIN KINASE.
 FT BINDING 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 587 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 124 124 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAc...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAc...) (POTENTIAL).
 FT VARIANT 654 654 I->V.
 FT VARIANT 655 655 /FTID=VAR_004077.
 FT CONFLICT 1170 1170 /FTID=VAR_004078.
 FT SEQUENCE 1255 AA; 137909 MW; 3989DFDA04C9F62 CRC64;
 SQ
 Query Match 96.9%; Score 6604; DB 1; Length 1255;
 Best Local Similarity 96.7%; Pred. No. 0;
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;
 QY 1 METALCRWGLLALLPQASTQVCTGDMKRLPASPEHLDMLRLYQCCQVVGNTL 60
 DB 1 MELALCRWGLLALLPQASTQVCTGDMKRLPASPEHLDMLRLYQCCQVVGNTL 60
 QY 61 ELTYLPTNASLFLDDIQEVQVYLIANQVRQVPLQRLRYRGQLTFEDNVALVDNG 120
 DB 61 ELTYLPTNASLFLDDIQEVQVYLIANQVRQVPLQRLRYRGQLTFEDNVALVDNG 120
 QY 121 DELNNTPTVTGASPGGLRELQRLSTELIKGVLQRPOLCYOPTILMKDIFHKNQLA 180
 DB 121 DELNNTPTVTGASPGGLRELQRLSTELIKGVLQRPOLCYOPTILMKDIFHKNQLA 180
 QY 121 DELNNTPTVTGASPGGLRELQRLSTELIKGVLQRPOLCYOPTILMKDIFHKNQLA 180
 DB 121 DELNNTPTVTGASPGGLRELQRLSTELIKGVLQRPOLCYOPTILMKDIFHKNQLA 180
 QY 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRYCAGGACRCAGPLTDCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRYCAGGACRCAGPLTDCHEQC 240
 QY 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRYCAGGACRCAGPLTDCHEQC 240
 DB 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRYCAGGACRCAGPLTDCHEQC 240
 QY 241 AAGCTGPRKSDCLALHFHNSGICELHCPALVTYNTDFESPMBEGRYTFASCVTTAC 300
 DB 241 AAGCTGPRKSDCLALHFHNSGICELHCPALVTYNTDFESPMBEGRYTFASCVTTAC 300
 QY 301 YVYLSIDVSGCTLVCPFLHNOEYTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
 DB 301 YVYLSIDVSGCTLVCPFLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHREVRATSAN 360
 QY 361 ELEFACCKKIFGSLAFIPSSFDGDPASNTAPLOPQLOVFEETLETGYLISAMPDSLP 420
 DB 361 IOEFACCKKIFGSLAFIPSSFDGDPASNTAPLOPQLOVFEETLETGYLISAMPDSLP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGSRSLRELSGALIHNNHLCFVTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLISWLGSRSLRELSGALIHNNHLCFVTV 480
 QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGACGQLCARGCMKPGPTQCNCGFLRGQEC 540
 DB 481 PMDQLFRNPHQALLHTANRPEDECVGEGACGQLCARGCMKPGPTQCNCGFLRGQEC 540
 QY 541 VEECRVLQGLPREYVYARHCLPCHPECCQPNQNSVTCFGEADQCAVCAHYKDPFCVAC 600
 DB 541 VEECRVLQGLPREYVYARHCLPCHPECCQPNQNSVTCFGEADQCAVCAHYKDPFCVAC 600
 QY 601 PSGVRDLSTYMPWKPEDEGAQCPCINCTHSVDLDDKCPAQRASPLTYSAYVG 660
 DB 601 PSGVRDLSTYMPWKPEDEGAQCPCINCTHSVDLDDKCPAQRASPLTYSAYVG 660
 QY 661 ILVVVLLGVVFGILI-----FNNFVSFWLVRPKVASAHLEPLTPSGAMPQAQVRIK 714
 DB 661 ILVVVLLGVVFGILI-----FNNFVSFWLVRPKVASAHLEPLTPSGAMPQAQVRIK 714
 QY 661 ILVVVLLGVVFGILI-----FNNFVSFWLVRPKVASAHLEPLTPSGAMPQAQVRIK 714
 DB 661 ILVVVLLGVVFGILI-----FNNFVSFWLVRPKVASAHLEPLTPSGAMPQAQVRIK 714

FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD_RSS 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RSS 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA, 138831 MW, 6129264583011402 CRC64;
 Query Match 85.4%; Score 5817; DB 1; Length 1257;
 Best Local Similarity 85.2%; Pred. No. 5,16-306;
 Matches 1075; Conservative 57; Mismatches 117; Indels 12; Gaps 4;
 1 MELALALCRWGLLALPPGASTQVCTGDMKRLPLASPEHMLDMLHYOGCQVQGNL 60
 1 MELAMCRWGLLALPPGASTQVCTGDMKRLPLASPEHMLDMLHYOGCQVQGNL 60
 61 ELTYLPTNASLFLQDIOEVQVYLIAHNOYQVPLQRLIVRGTOLEFEDNYALAVLDNG 120
 61 ELTYLPANASLFLQDIOEVQVYMLIAHNOYKAVPLQRLIVRGTOLEFEDNYALAVLDNR 120
 121 DPLNNTPTVT-GASPGGLRELUQASLTILKGSVLIQKRNPOLCYQDTILMKDIFHKNNQL 179
 121 DPOENVAASTPGRIPEGLRELQSLTEILKGVLIKRNPOLCYQDVMWLMKDVPRKNNQL 180
 180 ALLTIDTRSRACHPCSPMCKGSRMGSSSEDCOSLFTVTCAGGACARCKGPLPDCCHEQ 239
 181 APVIDIDTRSRACPCACACCKDNHMGSPEDCQILGICITSCACARCKGPLPDCCHEQ 240
 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMFNPBGRYTFEGASCYATC 299
 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMFNPBGRYTFEGASCYATC 300
 300 PYNVISTDVGSCITVCPHNOEVTAEDETCCEKSKPCAVCYGLGNOYVKAHSKFI 359
 301 PYNVISTEVSCTLVCPHNOEVTAEDETCCEKSKPCAVCYGLGMEHNRGARALTSD 360
 360 TELEFAGCKKIIFGSLAFIPESFDGPASNTAPLOPEQLOVEFTELEITGYLYISAMPDSL 419
 361 NVQEFDCCKIIFGSLAFIPESFDGPSSGIALPRLPEQLOVEFTELEITGYLYISAMPDSL 420
 420 PDLVYFQNLQVIRGILHNGAYSLTLOGIGTSMGLRSLRELSGLALIHNTHLCPVHT 479
 421 RDLVYFQNLRIIRGILHNGAYSLTLOGIGTSMGLRSLRELSGLALIHNTHLCPVHT 480
 480 VPMQDLFRNPHQALHTANRPEDC-CVGEGLACHOACARGHGWGPTQVCNCSQFLRQ 538
 481 VPMQDLFRNPHQALHTANRPEDC-CVGEGLACHOACARGHGWGPTQVCNCSQFLRQ 540
 539 ECVEECRYLQGLPREYVNAARHCLPCHPECOQNGSVLTFFGEPADOCVACAYKPPFYA 598
 541 ECVEECRYKGLPREYVNAARHCLPCHPECOQNGSVLTFFGEPADOCVACAYKPPFYA 600
 599 RCPGSKPDLSTYMPKPFDEEGACQPCPINTHSCVDLDBKGCFAEORASPLTSISAV 658
 601 RCPGSKPDLSTYMPKPFDEEGACQPCPINTHSCVDLDBKGCFAEORASPLTSISAV 660
 659 VGILLVVLGVFGILLI-----FNNFTVSEFWLRPKYSASLSEPLTSSGAMPNOAQMRI 712
 661 VGILLVVLGVFGILLI-----FNNFTVSEFWLRPKYSASLSEPLTSSGAMPNOAQMRI 716
 713 LKETEIRKVLGSGAFGVYKGINI PDGENVKIIVAIKVLRENSPANKEIIDEAAYM 772
 717 LKETEIRKVLGSGAFGVYKGINI PDGENVKIIVAIKVLRENSPANKEIIDEAAYM 776
 773 AGVSPVYVSRLLGICLTSTVQLVTOLMPYGLLDHVRNCRGLSGQDILNMCQIAKMS 832

Db 777 AGVSPVYVSRLLGICLTSTVQLVTOLMPYGLLDHVRNCRGLSGQDILNMCQIAKMS 836
 Qy 833 YLEEDVRLVARDLAARNVLYKSPBNHYKTFDGLARLLDIDEYHADGKVPKMALESI 892
 Db 837 YLEEDVRLVARDLAARNVLYKSPBNHYKTFDGLARLLDIDEYHADGKVPKMALESI 896
 Qy 893 LRRRPHOSDWSYGTATWELMTFGAKPVDGI PAEIPDLLEKGRLLPQPICTIDVYMI 952
 Db 897 LRRRPHOSDWSYGTATWELMTFGAKPVDGI PAEIPDLLEKGRLLPQPICTIDVYMI 956
 Qy 953 MYKCMWIDSECRPREFELVSEFSRMAPDQRFVVIQNEBLSGASPLDSTFYRSLLEDDM 1012
 Db 957 MYKCMWIDSECRPREFELVSEFSRMAPDQRFVVIQNEBLSGASPLDSTFYRSLLEDDM 1016
 Qy 1013 GDLVDAEYLVYQQGFCDPAPGAGMYHHHSSSTSGGGDLTLGLSEBEEPRSP 1072
 Db 1017 GDLVDAEYLVYQQGFCDPAPGAGMYHHHSSSTSGGGDLTLGLSEBEEPRSP 1076
 Qy 1073 LAPSEAGSDVFDGDLGMAAGKGLSLPFDHDSPLQRYSEDEPTVPLPSETDGVAPLTC 1132
 Db 1077 LAPSEAGSDVFDGDLGMAAGKGLSLPFDHDSPLQRYSEDEPTVPLPSETDGVAPLTC 1136
 Qy 1133 PQREYVNOQEVQPOPLPREGPLPAAPRAGATLEPAKTLSPGKNGYKDVAFGAVENP 1192
 Db 1137 PQREYVNOQEVQPOPLPREGPLPAAPRAGATLEPAKTLSPGKNGYKDVAFGAVENP 1196
 Qy 1193 EYLTPOGGAAPQHPAPRPAFPAFNDLYYWDODPPEBAGAPSPFKGPTANPEYGLDVP 1252
 Db 1197 EYLTPOGGAAPQHPAPRPAFPAFNDLYYWDODPPEBAGAPSPFKGPTANPEYGLDVP 1256
 Qy 1253 V 1253
 Db 1257 V 1257
 RESULT 3
 ERB2_MESAU STANDARD; PRT; 1254 AA.
 ID ERB2_MESAU
 AC 06053;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) [C-erbB-2].
 GN ERBB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 CX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.,
 RT Cloning and activation of the Syrian hamster neu proto-oncogene."
 RL Gene 140:251-255(1994).
 CC FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC SUBCELLULAR LOCATION: TYPE I membrane protein.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: D16295; BAA03801.1; -
 DR HSSP: P11362; IFCK
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001214; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYRCK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TDR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_TDM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transfease; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.

FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 158 368 CYS-RICH.
 FT DOMAIN 472 644 CYS-RICH.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.
 FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.1%; Score 5801.5; DB 1; Length 1254;
 Best Local Similarity 84.8%; Pred. No. 3.5e-305;
 Matches 1068; Conservative 66; Mismatches 114; Indels 11; Gaps 3;

QY 1 MELALCRWGLLLALPPGAASVGVCTGTDMKRLTPASPEHLLMLRHLYGCGVQGNL 60
 DB 1 MELAAKMGKLLALLBPASGASVGVCTGTDMKRLTPASPEHLLMLRHLYGCGVQGNL 60
 QY 61 ELTYLPNTASISFLQDIQEVGVYLLAHNVQVPLQRLIRVGTQQLFEDNYALALVLDNG 120
 DB 61 ELTYLPANATLSFLQDIQEVGVYLLAHNSQVRFHPQLRLIRVGTQQLFEDNYALALVLDNR 120
 QY 121 DPLNNTTPVNGASFGRIRELOLRSTELILKGVYLIGNNPOLCYQDITLMKDIFFKNNOLA 180
 DB 121 DPLNNTVATGRTEGRIRELOLRSTELILKGVYLIRGNPOLCYQDITLMKDVFRKNNOLA 180
 QY 181 LTLIDNRSRACHPCSPMKCKSRQCGESSEDDQSLTRTVAGACGACRKGPLPTDCHEQC 240
 DB 181 PVDIDNRSRACHPCPCAPACKONHCWASPEDCQTLTGITAPRAVAPARARLPTDCHEQC 240
 QY 241 AAGCTGPKGSDCLALFHNHSGTELCALPALVYNTDTFESMPNPEGRYTFGASCVTTCP 300
 DB 241 AAGCTGPKGSDCLALFHNHSGTELCALPALVYNTDTFESMPNPEGRYTFGASCVTTCP 300
 QY 301 YNYLSTVGSCTTLCPLHNOEVTAEDETOGCEKSKPCARCYGLGMQYIKANSKFTIGT 360
 DB 301 YNYLSTVGSCTTLCPLHNOEVTAEDETOGCEKSKPCARCYGLGMHILGARAITSAN 360
 QY 361 ELEPAGCKKIFGSLAFIPESFSDGPASNTAPLOEQLQVETLEITGYLISAMPDLP 420
 DB 361 IQEPAGCKKIFGSLAFIPESFSDGNSSGIALPTEQQLVETLEITGYLISAMPDLP 420
 QY 421 DLSVFNQVIRGRILNNGASLTLQGLISWGLRSLRELQSLALIHNTHLCFVHTV 480
 DB 421 DLSVFNQVIRGRILNNGASLTLQGLIRMLGRSLRELQSLALIHNTHLCFVHTV 480
 QY 481 PMDQLFRNPHQALLHTANRPEDECVGGLACHOLCARHCWGPPTQCVNCSQPLRQEC 540
 DB 481 PMDQLFRNPHQALLHSGNPSSEECGLKDPACYPICAHGHCWGPPTQCVNCSHPLRQEC 540
 QY 541 VEEGRVLOGLPREVYNAHCLPCHPEQPOQNGSVTCGPEPDOCAVCAHYDDPPFCVARC 600
 DB 541 VKECRVWGLPREVYNGHCLPCHPEQPOQNGSVTCGSEADQCTACVHYDPSFCVARC 600
 QY 601 PSGVYKPDLSYPIKPFDEEGACQPCINCHSCVDDDDKCPAEGPASPSTISVSAVAG 660
 DB 601 PSGVYKPDLSYPIKPFDEEGACQPCINCHSCVDDDERCPAEGPASPSTISVSAVAG 660
 QY 661 ILTVVYVGLVGVGIIIL-----FNNFTYSFWLRVYKVSASHLEPLTPSGAMPNQOMRLK 714
 DB 661 ILTVVYVGVVGVGIIILKRRQKIRKYTWKRLQETEL-----VEPLTPSGAMPNQOMRLK 714
 QY 715 ETELRYKVVGLSGAFGTYYKGIWIPDGENYKIPVAIKYLRNTSPKANKELDFAVYVAG 774
 DB 715 ETELRYKVVGLSGAFGTYYKGIWIPDGENYKIPVAIKYLRNTSPKANKELDFAVYVAG 774
 QY 774 VGSPPVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRGRGLSGODLLNMCQIAKMSYL 834
 DB 774 LGSPPVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRGRGLSGODLLNMCQIAKMSYL 834
 QY 835 EDVLLVHEDLAARVNVVSPHVKITPQGLARLLDIDETEHNAQGVKIPMMLESTLR 894
 DB 835 EDVLLVHEDLAARVNVVSPHVKITPQGLARLLDIDETEHNAQGVKIPMMLESTLR 894
 QY 895 RRFTHQSDVMSYGVTVLWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMV 954
 DB 895 RRFTHQSDVMSYGVTVLWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMV 954
 QY 955 KCMWIDSCRRFRFELVSEFSRMARDQRFPVVIQNEDLGPASPLDSTFYRLLDDDDMGD 1014
 DB 955 KCMWIDSCRRFRFELVSEFSRMARDQRFPVVIQNEDLGPASPLDSTFYRLLDDDDMGD 1014
 QY 1015 LVDAEEYLVPPQGFCEPDPAAGAGVYHRRSSSTSGCGDLTLGLEPSEEAAPRSPLA 1074
 DB 1015 LVDAEEYLVPPQGFCEPDPAAGAGVYHRRSSSTSGCGDLTLGLEPSEEAAPRSPLA 1074

QY 1075 PSEAGSDVFDGDIKMGANGLOSLFTHDPSLQRYSEDTVPJPSSTGCVUAPLTCSPQ 1134
 DB 1077 PSEAGSDVFEGETGKATGKPOSISPRDLSPLQRYSEDTLPJPTETDGVAPLACSPQ 1136
 QY 1135 PEYVNOPDVPPOPSPEEGPLPARPAGATLERAKTSLPKNGVXKQVAFGGAVENPEY 1194
 DB 1137 PEYVNOFEVRPQPLTEGEPFVPRPAGATLERKTSLPKNGVXKQVFTFGAVENPEY 1196
 QY 1195 LTPOGGAPOPHPPAPFPAFNDLYWDQDPPEKADPSFTFKGTPTAENPEYLGDDVP 1253
 DB 1197 LVPRGSGASQPH-PPALCPAFDNLXYWDQDPSEKSGSPNTEGTPAENPEYLGDDVP 1254

RESULT 4
 EGFR_HUMAN STANDARD; PRT; 1210 AA.
 AC P00533; P06268; Q14225; Q9JMD7; Q9JMD8; Q9JMG5; Q92795; O00732;
 AC O00688; Q9B252; Q9H2C9; Q9GZK1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase EDB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6128312;
 RA Ullrich A., Cossens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Mittels N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=95382957; PubMed=7654368;
 RA Tlekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=97076686; PubMed=8918811;
 RA Reiter J.L., Mainle N.U.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=97256547; PubMed=9103388;
 RA Tlekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX MEDLINE=21100872; PubMed=1161793;
 RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Scheel Sincilar C., Pearall R.S., Green P.J., Yee D., Lampiad A.L.,
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Mainle N.U.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20(2001).

RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Reiter J.L., Thredgill D.W., Danielsen A.J., Scheel C.M.,
 RA Lampiad A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Mainle N.U.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor.";
 RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Kruliger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGFR receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Stimmann F.A., Gepe M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217353; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kitchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Wosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honnegger A.M.,
 RA Howk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)

RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor."
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=66398132; Pubmed=6962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts."
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; Pubmed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor."
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; Pubmed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Iax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor."
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; Pubmed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens."
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL: X00588; CA25240.1; -
 CC EMBL: U05089; AAB53063.1; -
 CC EMBL: U48722; AAC50802.1; -
 CC EMBL: U48723; AAC50804.1; -
 CC EMBL: U48724; AAC50796.1; -
 CC EMBL: U48725; AAC50797.1; -
 CC EMBL: U48726; AAC50798.1; -

Query Match 45.0%; Score 3067; DB 1; Length 1210;
 Best Local Similarity 49.1%; Pred. No. 6.3e-158;
 Matches 621; Conservative 174; Mismatches 362; Indels 108; Gaps 24;

QY 11 LLLALLPGEAA--STVCTGTDMKRLPASPETHLDMRHLYOGGVVQGNLELYLPTN 68
 DB LLLAALCPASRLAEKKVCGTSNKLTLQGTEDFDHLSLQRMNNEVVLGNLEIYVQRN 73
 QY 69 ASLSFLQDIQEVGVYVLIAPHNOVCPLORLRIYVGTQLFEDNVALVAVLNDGDPNNTPE 128
 DB YDLSFLKTIQEVAGVYVLIAPLNTVERIPLENLIQIIGNNYENSVALAVLSND----- 126
 QY 129 VTGASPGGLRELOLSLREILKGGVLIQRNPOLQODTILIMMDIHHKNNQALALLIDTNR 188
 DB ---ANKTGKLELPMNLQELILGAVRFSNNPLQVESIQMRDISSDFLSMMSDFQNH 183
 QY 189 SRACHPCSPMKGRSBRCSWESSEDCSLTRVYAGGCA-RCKGPLPTDCCEGCAAGCTGP 247
 DB LGSCKQCDPSRCNGSCWGEENCCOKLTKITCAQCCSRCRCKSPSCCHNCAAGCTGP 243
 QY 248 KNSDCLACHFHSGLCELCALVTYNTDTRESNPNREGRTFGASCVTACPYVYLSLD 307
 DB RSDDLVCRRKFEDEATCKDTCPLMLYPTTYQMDVNPBGKYSFGATCVKCPRYVYVTD 303
 QY 308 VGSCTLVCPLNHQTAEQGTQRCCEKCKPCARVCYGLQMYIKANSKFIGITELE-FAG 366
 DB HSCVTRACGADSYEM-EDGVYKCKCKCGPCCKVNGGIGIEFK-DLSLNATNTNKHFNQ 361
 QY 367 CKKIFGSLAFLESFSDGPASNTAPLOPQLOVFTLEITGYLYISAMPDSLPLSYVQ 426
 DB CTSISGDHLILFVAFRGDSFTHTPDLQEDLILTKVKEIFGLIQAMPENRDLHAFE 421
 QY 427 NLOVRGRILHNGAASLTLOGIGISMLGRSLREISGLALHNHTHLCFYVTVMDOLE 486
 DB NLEIRGRKQKQGSFLAVSLNITSLGRSLKESIDGVIISGNKLCYATNTVMKCLF 481
 QY 487 RNPQALLHTANRPDECEVGEGLACHQDCAHGCWGPPTQCVNSQPLRGOECVEECV 546
 DB GTSQCKTKLISNRGENSCATQVCHALCSPEGCWGPBRDVCSSRNNSRGECVCKNL 541
 QY 547 LOGLREYVNAHCLPCHPECCPQNGSVTCFEPENDQCVACHYKADPPCVARCGSVKPE 606
 DB LGEGEREPFENSECIQCHCECLPQAMNITTCRGPDNCTICQAHYIDGPVCVTCAGVWG 601
 QY 607 DLSYPIWKFPEDEBACQCPPINCHSCVDLDDCKPAEQRASPLTSIVSAVG--IIL 663
 DB ENNTL-VKMYAAGVCHLCHNCYTCGTCGPELBCPTNGKRP--STATGVGALLLL 658
 QY 664 VVLLGVVGLILFNPNFTV-SFWLVYKVSASHLEPLTPSGAMPNOQMILKELRLVKY 722
 DB VVALGI--GLFMRHRHIVKRTLRLLQERELVEPLTPSGEAPNOLLRLIKETEFKKIK 716
 QY 723 VGSAGFGVYVYGMIPDQENVKIPVATKVLRENTSPKANKXILDEAVYMAVGSPVYSR 782
 DB VLSGAGFVYVYGLMIPDEKVKITVALKELEKREATSPKANKXILDEAVYMASVDNPHCR 776
 QY 783 ILGICLTSTVQVLTQIMPYGCLLDHVRNRLGSGQDLLMCMQIAKGSYLEDEVLYHR 842
 DB LIGICLTSTVQVLTQIMPYGCLLDHVRNRLGSGQDLLMCMQIAKGSYLEDEVLYHR 836
 QY 843 DLAANNVLYKSNHKKINDPGLARLDIDETFNADGKVPKMMALSLIRRTTHSD 902
 DB DLAANNVLYKTPQHKITIDFGLAKLGAEEKEYHAGGKVPKMMALSLIRRTTHSD 896
 QY 903 VMSYGVTVLWELTFGAKPYDGIIPAREIDPLEKSGRLPOPICTIDYVIMVYKCMNIOSE 962
 DB VMSYGVTVLWELTFGAKPYDGIIPAREIDPLEKSGRLPOPICTIDYVIMVYKCMNIOSE 956
 QY 963 CRRPRELVSEFSKMAKDPQRFVYVQ-NEDLCFASPPLDSTFRSLLEDNDGDLVDAEY 1021
 DB SRKPRELTIIFSMAKDPQRFVYVQ-NEDLCFASPPLDSTFRSLLEDNDGDLVDAEY 1016
 QY 1022 LVPGQGFPCPDPAAGGVVHHRHRSSTRSGGDLTLGLEPSESEAPRSPLAPEGAGS 1081
 DB LVPGQGFPCPDPAAGGVVHHRHRSSTRSGGDLTLGLEPSESEAPRSPLAPEGAGS 1042

FT DISULFID 620 628 BY SIMILARITY.
 FT DISULFID 624 636 BY SIMILARITY.
 FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL).
 FT CONFLICT 19 19 C -> S (IN REF. 2).
 FT CONFLICT 539 539 C -> W (IN REF. 5).
 FT CONFLICT 991 991 L -> F (IN REF. 4).
 FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
 SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match 44.6%; Score 3041; DB 1; Length 1210;
 Best Local Similarity 48.9%; Pred. No. 1,6e-156; Indels 114; Gaps 25;
 Matches 622; Conservative 168; Mismatches 369;

11 LLLALLPPGAA--STGYCTGTDMKRLPASPETHLDMLRLHLYOGCCOVQGNLELYLPTN 68
 14 LLLALLPPGAA--STGYCTGTDMKRLPASPETHLDMLRLHLYOGCCOVQGNLELYLPTN 73
 69 AASSPLDIDIEVQGYVILANNOVROVLORLATVYRGQLFEDVYALAVDNGPPLNNTTP 128
 74 YDISFKTIDIEVQGYVILANNOVROVLORLATVYRGQLFEDVYALAVDNGPPLNNTTP 124
 129 VTGASPGGLRELGLRLSLTEILKGVILQRPOLCYOCTILMKDI---FHKNQALATLI 184
 125 -YGTNRGLRELGLRLSLTEILKGVILQRPOLCYOCTILMKDI---FHKNQALATLI 180
 185 DTNRSRACHPGSPCKSPKSCRCWGSSEDDQSLRTVCAQGA-RCKAPLPDCCHEQCAAG 243
 181 -QSHPSGCPKCDSPCKSPKSCRCWGSSEDDQSLRTVCAQGA-RCKAPLPDCCHEQCAAG 239
 244 CTGPKASDCLACIHFHNSG:CELCAPALVTYNTDFESMNPGRGTFGASCYACAPVNY 303
 240 CTGPKASDCLACIHFHNSG:CELCAPALVTYNTDFESMNPGRGTFGASCYACAPVNY 299
 304 LSTDVGSCTLVCPILHNOEVTAEQGTORCEKSKPCARVCGYGLMOYIKANSKEFIGITELE 363
 300 VVMDHSGCVACGPDYEV-EEDGIRKCKKCDGPRCKVCGYGLMOYIKANSKEFIGITELE 357
 364 -FAGCKKIFGSLAFLESPFDGPDASNTAPLQPSQLOVETLEITGTYLISANPSLPLD 422
 358 HFYKCTAISGDLHILPAFKGDSPTPTPLDPELELLKTKVTKITGTYLISANPSLPLD 417
 423 SVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRELSSGALITHNTLGFVTPW 482
 418 HAFENLEIRGRILHNGAVSLTQGLISWLGRLSRELSSGALITHNTLGFVTPW 477
 483 DQFRNPHOALLTANRPEDECVGEGILACHQLCARHGCHWPGTQCVNCSQFLRGEQV 542
 478 KKLFGTPNOKTKIMNRAEKDCAVHVNCPNLSGSCWGPREDVCSQNVSRGREQV 537
 543 EGVVLQGLPREYVNAHCLPCHPECQFONGSVTCFGEADQVACAHYDPPPCVAPCS 602
 538 KNILTEGEPRERFENSCICGHECLPQANNTICTRGPNCTGCHAYIDGPRCVATCA 597
 603 GVPADLSTYPMKFPDEGACQCPINCHSCVDLDDKGPAPQASPLTSIYSAVAVIL 662
 598 GINGENNNTL-VMKYADANNVCHCHANCYIGCAGPGJGCEVWPSGPKIPSIATIGVGL 656
 663 LVVVLGVVFGILIF--NNFTVSFWLVRPKVASHLPELTPSGAMPNQMRILKETEIR 719

DB 657 LFIIV-VAGIGLFLRRRIIVAKRTLRLLQRELEVEPLTSGEAPNOAHILKETEER 715
 QY 720 KYKULGSAFGVYVGMIPROSEANKIVAVIUVLEBNSPKANKELILEAYVMAGVSPY 779
 DB 716 KIKVIGSGAFGVYVGMIPSEKIVALELEBNSPKANKELILEAYVMAGVSPY 775
 QY 780 VSRLGICLSTVQVLTQVLMQYGLLDHRENRGLSGODLLNMCQIAKMSYLEDVL 839
 DB 776 VORLLGICLSTVQVLTQVLMQYGLLDHRENRGLSGODLLNMCQIAKMSYLEDVL 835
 QY 840 VHRDLAARVLYKSNHVKITDFGLARLLDIDETVHADGCVPIKMMALBETLRRFTH 899
 DB 836 VHRDLAARVLYKSNHVKITDFGLARLLDIDETVHADGCVPIKMMALBETLRRFTH 895
 QY 900 QSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLKGERLPQPICTIDVYIMYKMMI 959
 DB 896 QSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLKGERLPQPICTIDVYIMYKMMI 955
 QY 960 DSECRPRRELVSSEFSRMAPDQRFVITQ-NEDLGASPLDSTFYRSLLDDMGDLVDA 1018
 DB 956 DADSRPKRELILLESKVARDPQRYLVIQGDERMHLPEPDTSNFYRALMEDMEDVYDA 1015
 QY 1019 EEYLPQCGFFCPDPAFAGGCVHHRHSSSTRSGGDLTLGLESSEEARPSPLAPSEG 1078
 DB 1016 DEYLPQCGFFCPDPAFAGGCVHHRHSSSTRSGGDLTLGLESSEEARPSPLAPSEG 1041
 QY 1079 AGSDVFDGDLGMAAGLQSLPETHDPSPLQRYSEDPVLPSET--DGYVAPLTCSPQPE 1136
 DB 1042 ATSN-----NSTVACINRSGSCVKEKDAFLQRRSSDPFGAVYEDNIDDAFL-----PYPE 1091
 QY 1137 YNQGDPVQPPSPREGPLPAPRPGATILERAKTLSPOKGVYKQVFAFGAVENPEYL- 1195
 DB 1092 YNQGDPVQPPSPREGPLPAPRPGATILERAKTLSPOKGVYKQVFAFGAVENPEYL- 1140
 QY 1196 TPQGAAPQPPSPREGPLPAPRPGATILERAKTLSPOKGVYKQVFAFGAVENPEYL- 1239
 DB 1141 TQO-----PTLSSGFSNPAIMQKSHQSLNDPDIQODFPFKETKNGIFKG-P 1190
 QY 1240 TAENPEYGLDVP 1252
 DB 1191 TAENPEYGLDVP 1203

RESULT 6
 ERB4_HUMAN
 ID ERB4_HUMAN STANDARD; PRT; 1308 AA.
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p106erbB4) (tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=93189574; PubMed=8383326;
 RA Plozman G.D., Couloucou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 Foy L., Neubauer M.G., Shoyab M.,
 "Ligand-specific activation of HER4/p106erbB4, a fourth member of the
 RT epidermal growth factor receptor family,"
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).
 RC TISSUE=fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,
 Klagsbrun M.;

[illegible]

```

Db      476 TLFSTINQIVIRDRNRKAENCTAGBVCNHLSSDGCWGPDPQCLSCRRFSRGRICIES 535
Qy      544 CAVLGLPREVYNAHCHLPCHEPCOP-QNGSVTCGPREDDQCAAHYNDPPFCARCS 602
Db      536 CMLYGEFFREFENSGICVBCDPOCEMDEGLTCHGPGPDNCTKCSHFADGPNCEKCD 595
Qy      603 GVKPDLSTYPMIKWKFDEBACQPCPCINTCHSCVDDDKCC-----PAEORASPL 651
Db      596 GAGANSF--IFRYADPDRECHCHCHNCTGCGNCPSTSHDCIYYPWTHSTLQGHAR-TPL 652
Qy      652 TSIVSAVY-GILLVVLGVVPGILLFNNFTVSFWLRVPVNSH-----LEPITBSGA 703
Db      653 --IAAGVIGLFLVLVLGLTFAVYARR-----SIKKRLRRLFELETELVEPLTPSGT 703
Qy      704 MENOAMRIKETELRKVKVLSGAFYVKGIMIPDGENVKIPIAKVLRNTPKANK 763
Db      704 ANQAGRLIKETELKRVKLVLSGAFYVKGIMIPDGENVKIPIAKVLRNTPKANK 763
Qy      764 ELIDEAVVAVGVSPPVSRLLGICLTSTVQVLTQMLPYGCLLDHYRENRGLSGODLLM 823
Db      764 EEMDEALIMASMDHPHLLVLLGLVCLSPITQVLTQMLPHGCLLEVYHEHKNIGSOLLNM 823
Qy      824 CMOIAGMSYLEDVRLVNRDLAARVLYKSPNHVKITDGLARLLDIDETEHADGKVP 883
Db      824 CMOIAGMSYLEERRLVNRDLAARVLYKSPNHVKITDGLARLLDIDETEHADGKVP 883
Qy      884 IKMALESILRRRFTHQSDVWSYGVTWELMTFGAKPYDGPAREIPDLLEKGERLPQPP 943
Db      884 IKMALECIHYRKFTHQSDVWSYGVTWELMTFGAKPYDGPAREIPDLLEKGERLPQPP 943
Qy      944 ICTIDVYVMWCMWTDSECPREFRLVSEFSRMAPDPRFVIONED-LGASPLDSTF 1002
Db      944 ICTIDVYVMWCMWTDSECPREFRLVSEFSRMAPDPRFVIONED-LGASPLDSTF 1002
Qy      1003 YRSLLDEDDMCDLVDAEVLVPOQGFCCDPAPAGAGWVHRRSSSTSGGDDTLGLE 1062
Db      1004 FQNLDEDELDMMQDEYLVF-QAENIPPP-----IYTSARLDSNKS-----EIGHS 1051
Qy      1063 PSEEEAPRS-----PLAP-SEGAGSVFPGDGLGMAA 1093
Db      1052 PPAYATPMGNGFVYRDGFAAEQGVSVYRAPSTIPAPVAGATAIIFDSCNGTL 1111
Qy      1094 KGLQSLPTHDPSLPQRYSEDPVPLPS-----ETDGYVALPCTSPQPEYVNOQDVPRQ 1146
Db      1112 KRPVAPVHVEDSSTQYISADPIYFAPERSPRGELDEGIMTMRDKPKQEYLNPEV--- 1167
Qy      1147 PPSPRSGPLPAARPAATLERAKTISPKNVGVKDYFAFGAVENPEYLTPOGGAAPQH 1206
Db      1168 -----ENPFVSR--KMGDLQ-----ALDPEYHNASNG----- 1194
Qy      1207 PPPA-----FSPAFLNLYVWDQDPPRGA-- 1230
Db      1195 PPKADEYVNEPLYLNTFANTLGKAEYLNKILNSPEKAKKAFDNPDIWNSHLSPPRSTLQ 1254
Qy      1231 PSTFGKPTT-----AENPEYL 1247
Db      1255 HPDYLOEYSTKYFYKONGRIRPIVAVENPEYL 1285

```

```

Rp      SEQUENCE FROM N.A.
Rc      TISSUE=Heart;
Rc      MEDLINE=98221155; PubMed=9553078;
Ra      Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
Ra      Marchionni M.A., Kelly R.A.;
Rt      "Neuregulins promote survival and growth of cardiac myocytes.
Rt      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
Rt      ventricular myocytes."
Rl      J. Biol. Chem. 273:10261-10269(1998).
Rn      [2]
Rp      SEQUENCE OF 848-901 FROM N.A.
Rc      TISSUE=Sciatic nerve;
Rc      MEDLINE=9122560; PubMed=2025425;
Ra      Lai C., Lemke G.;
Rt      "An extended family of protein-tyrosine kinase genes differentially
Rt      expressed in the vertebrate nervous system."
Rl      Neuron 6:691-704(1991).
Rn      [3]
Rp      SEQUENCE OF 1031-1198 FROM N.A.
Rc      STRAIN=Sprague-Dawley; Tissue=Spinal cord;
Rc      MEDLINE=97184212; PubMed=9030624;
Ra      Carroll S.L., Miller M.L., Frohneert P.W., Kim S.S., Corbett J.A.;
Rt      "Expression of neuregulins and their putative receptors, ErbB2 and
Rt      ErbB3, is induced during Wallerian degeneration."
Rl      J. Neurosci. 17:1642-1659(1997).
Rn      [4]
Rp      FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
Rp      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BTACELULIN AND
Rp      NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
Rp      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
Rp      CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
Rp      tyrosine phosphate.
Rp      SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
Rp      RECEPTORS (POTENTIAL).
Rp      SUBCELLULAR LOCATION: Type I membrane protein.
Rp      TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
Rp      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
Rp      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
Rp      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
Rp      HEART.
Rp      -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
Rp      RESIDUES (BY SIMILARITY).
Rp      -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
Rp      -----
Rp      This SWISS-PROT entry is copyright. It is produced through a collaboration
Rp      between the Swiss Institute of Bioinformatics and the EMBL outstation -
Rp      the European Bioinformatics Institute. There are no restrictions on its
Rp      use by non-profit institutions as long as its content is in no way
Rp      modified and this statement is not removed. Usage by and for commercial
Rp      entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
Rp      or send an email to license@ebi.ac.uk).
Rp      -----
Rp      EMBL, AF041938; AAD08899.1;
Rp      EMBL, U52531; AAC3051.1;
Rp      HSGP, P11362; IFGK.
Rp      InterPro: IPR000494; EGFR_L domain.
Rp      InterPro: IPR000719; Euk_DKinase.
Rp      InterPro: IPR002174; Furin-like.
Rp      InterPro: IPR001245; Tyr_Kinase.
Rp      InterPro: IPR004019; YLP_motif.
Rp      Pfam: PF00757; Furin-like; 1.
Rp      Pfam: PF00069; Kinase; 1.
Rp      Pfam: PF01030; Recep_L_domain; 2.
Rp      Pfam: PF02757; YLP; 2.
Rp      PRINTS: PR00109; TYRKINASE.
Rp      PRODOM: PD000001; Euk_DKinase; 1.
Rp      SMART: SM00261; FU; 4.
Rp      SMART: SM00219; Tyrc; 1.
Rp      PROSITE: PS00107; PROTEIN KINASE ATP; 1.
Rp      PROSITE: PS0011; PROTEIN KINASE DOM; 1.
Rp      PROSITE: PS00109; PROTEIN KINASE TYR; 1.
Rp      Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
Rp      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
Rp      SIGNAL 1 25 POTENTIAL.

```

FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ENRB-4.
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 652 675 POTENTIAL.
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 166 334 CYS-RICH.
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.
 FT NP_BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843 BY SIMILARITY.
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 225 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 138 138 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).
 FT CONFLICT 1062 1062 S -> N (IN REF. 3).
 FT CONFLICT 1080 1082 PRT -> SYR (IN REF. 3).
 SO SEQUENCE 1308 AA; 146957 MW; D944B8096A0B41 CRC64;

Query Match 42.4%; Score 2889; DB 1; Length 1308;
 Best Local Similarity 44.2%; Pred. No. 2,66-148;
 Matches 598; Conservative 189; Mismatches 353; Indels 172; Gaps 28;

QY 297 TACRYNYLSTDVSGCTLVCPFLHNOEYTAEDGTORCEKSKRACAVCYGLMOKYIKANSKF 356
 DB 291 KCPHNHV-VDSSSCVACACSSKREV-EENGIMCKRCTDIOCRACGIGITGSLMAQTV 348
 QY 357 IGITELEFACCKKIFGSLAFLESPFODGPAASNTAPLOELOYFETLEITGVLYISAMP 416
 DB 349 DSNWIDFINCTKINGNLIFLVGIGHDPYNAIDADPEKXNFRTRYREITGFPLNTQTPW 408
 QY 417 DSLPDLVFONQVYIRIRIILHNGAVSLTLQGLISWGLSLRELGSLLIHNTLFCF 476
 DB 409 PNMTDFSVSLVTGGVLYSGSLTLKQOGLTSLOFOSLEKISGNITDNSLVCY 468
 QY 477 VHTPMDOLFNPQHALLHTANPREDECEVGEGLACHOLCARHCWGFQPCVANCQFPLR 536
 DB 469 YHINMTLTFSTVQRIVIDNRRAENCTAEGMVCNHLGNDGCGGPGDCLSCRRFSR 528
 QY 537 GQCEVEEERVYQGLPREVYANRHLPGHPECQF-QNCSYTCFGEADQCVACAYKDPF 595
 DB 529 GKICIESCNLYDGEFFREENGISICVECDSCQEKMEDGLITGHGPNCTKCSHFKQGN 588
 QY 596 CVARCPGKPDLSYMPILWKEPDEGACQPCPINCTHSCVDLDDKGC-----PA 644
 DB 589 CVEKCPDVLQGANF--IFKXADQDRCHPCHPCTQCGNGPTSHDCIYPMWGTSLPQ 646
 QY 645 EORASPLTIVSAVV-GILLVTVLGVVFGILIFNNFTVSEFWLRPKYSASH-----LE 696
 DB 647 HAR-TPL--IAAGVIGGLFVLIVALTFAVYVRK-----SIKKRALRPLETELVE 696
 QY 697 PLTSGAMPNOAQRILIKETELRKVKVLSGAFGVYKGIWIPGENVKIPIVAIKVLREN 756
 DB 697 PLTSGAMPNOAQRILIKETELRKVKVLSGAFGVYKGIWIPGENVKIPIVAIKVLREN 756
 QY 757 TSPKANKEILDEAVYMAVGSPVYSRLIGICLTSTVOLVTLQMPYGLLDHVENRGLG 816
 DB 757 TSPKANKEILDEAVYMAVGSPVYSRLIGICLTSTVOLVTLQMPYGLLDHVENRGLG 816
 QY 817 SODLLNMCQIAKAGSVYEDVRIYARDAAANVLYKSPNHYKIDFGARLLDIDETEYH 876
 DB 817 SODLLNMCQIAKAGSVYEDVRIYARDAAANVLYKSPNHYKIDFGARLLDIDETEYH 876
 QY 877 ADGGKVPILKMMALLESILRRFTHSDVSYGVYTWELMTFAKPYDGI PAEIPDLLEK 936
 DB 877 ADGGKVPILKMMALLESILRRFTHSDVSYGVYTWELMTFAKPYDGI PAEIPDLLEK 936
 QY 937 ERLPQPPICITDVIYKIMYKCMWIDSECPREELVSEFSRMAKDPQRFVIONED-LGPA 995
 DB 937 ERLPQPPICITDVIYKIMYKCMWIDSECPREELVSEFSRMAKDPQRFVIONED-LGPA 995
 QY 996 SPLDSTFYRSLLEDMDMGLVDAEYVLPQGGFCPCPD----- 1033
 DB 997 SBNDSKFPONLDEEDLEDMDAEYVLP-QAFNIPPIIYSTRHIDNSRSEIGHSPPA 1055
 QY 1034 -APGAGGVNHRHRSSSTRSGGGLTLGLEPSEEAAPSPPLAPSGAGSDVDGDLGCA 1092
 DB 1056 YTPMSGGQFVYVODGGAFTDQG---NPMRYTATITSTIPAPVA--QGAATAMPDSDCCNGT 1110
 QY 1093 AKGLDLSLPHDSPLORXYSEDPVLPFS-----ETDQYVFLTCSPOPEVYNQDVRP 1145
 DB 1111 LKRPVPHVQEDSSFORISADPTVAPAEKNPAELEDDEGCMYPMHDKRQVELNVE--- 1167
 QY 1146 QPSPREGRPLAPARAGATLERAKTSLSPKNGVXKDVAFGAGAVENPEYLTPOGGAAPOP 1205
 DB 1168 -----ENPFVSR-----KNNDLO-----ALNPEVHSSASG----- 1194
 QY 1206 HPPPA-----FSPAFNLVYWDQDPERGA- 1230
 DB 1195 -PPKAEDEVYNEPLVINTTALGNAEYKNSLSVPEKAKAFNDPDMNSLPPRSTL 1253
 QY 1231 -PPSTFKGTP-----AENPEYL 1247
 DB 1254 QHPDYLAQEVSTKVFYKQNGRIRPIVAENEYLL 1285

RESULT 8
XMRK_XIPMA STANDARD; PRT; 1167 AA.

AC P1388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TU.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_Taxid=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Witbrodt J., Adam D., Malteschek B., Mauelel W., Raulf F.,
RA Telling A., Robertson S.M., Schartl M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
inducing Tu locus in Xiphophorus."
RL Nature 341:415-421(1989).
RN [2]
RP REVISION TO 515.
RA Schartl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL: X16891; CAA34770.2; -
DR PIR: S06142; S06142.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000179; Euk_kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00765; Kinase_1.
DR Pfam: PF01030; Furin-like_1.
DR Pfam: PF01030; Recep_Ldomain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00261; FU; 5.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00219; TYK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
KINASE
EXTRACELLULAR (POTENTIAL).
FT DOMAIN 26 642
TRANSMEM 643 665
FT DOMAIN 666 1167 POTENTIAL.
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 710 977
NP BIND 716 724 PROTEIN KINASE.
FT BINDING 743 743 ATP (BY SIMILARITY).
FT ACT_SITE 835 835 ATP (BY SIMILARITY).
BY SIMILARITY.

FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT DISULFID 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.4%; Score 2614.5; DB 1; Length 1167;
Best Local Similarity 44.2%; Pred. No. 1.4e-133;
Matches 562; Conservative 168; Mismatches 394; Indels 147; Gaps 30;

4 AALCWGLLLALPPGAST---QVCTGTMKRLRPSPEHLDMLRHLYOGCVOGN 59
8 AALLQ--LLVLVISIRCCSTDPDRKVCQGSINQM--LDNHLYKMKMISGCVALEN 62
60 LETLTPNASTSPLODIOEVQGYVLIHNOVRCVPLQRLRIVRGTOLEFEDNYALAVDN 119
63 LEITYGENODLSFQISQEVGYVLIAMNEVSTIPLVNLLIRGNLYEGNFTLLWSN 122
120 GDLPLNTPVTVGASGGRLREIQLNSLTILKGVLIQNPOLCYODTILMKDIFPKNQ 179
123 YQK-NPSSP--DYQVQGLQQLSLTLEILSGVYVSHNPPLCNVETINMMDIVKTSNP 179
180 ALLIDTNRSPRCHPSCPMCKGSRGSSSDCCSLTRTVAGGC-ARCKGPLPTDCCH 238
180 TMLILPHAFERQCKCHGCVNNGSCMAFGPHQCKFTLLCAEQCNRCRGKPIDCCNE 239
239 QCAAGCTGPKHSDCLAHFNHSGICEIHCALVYNTDTESMPNPRGRYTFGASCYTA 298
240 HCAGGCTGPRACTDCLACRDFNDGCTCKPTPYDIDVSHQVVDNPNIKYTFGAACVKE 299
299 CPYNYLSDVGSCTLVCLHNGEVAETGRCCEKSCPCARVCYGLGMOYIKANSKFG 358
300 CPSNTVATE-GACVRSAGSAGMLVD-ENGKSKCKDVCPCVCGIGSI-SVTIVN 356
359 ITEL-EFAGCKKIFGSLAFDPESPDGDASTAPLQAPOLQVFEITLBITGYLYISAMP 417
357 STNHSFNSCKTKINDIILNNSFEGDHYKIGTMDPHLWNLTTVEKITGYLYIMWPE 416
418 SLPLDSVFNQLVIRGRLIHNGAYS-LTLOGISWLGRLSRLREGSLALIHNTLCLF 476
417 NMTSLSVQNLEIRGRITFGSGFVVAVVQHLQWILRLSKVSAENVILKNTLQRY 476
477 VATVPMDFRNPFOALHTANRDECVGEGALCHQCARGHGCGPPTOCVNCSPCLR 536
477 ANTIMWRLLFSEDSIEYDART-----ENQCNKCSGDCGPPPTCVSGSLAHDR 529
537 GQECVEECRVLQGLPREYVNAHCLPCHPGCPQPNQSGVTCGPPADQCAAHYKDPFC 596

Db 530 GGCACVASCNLLQEPREAVDGRVCHQCELVDTSLTYGPGPANCSCSAFFQDQPC 589
 Qy 597 VAPCPGKPDLSYMPIMKPEDEGACQPCPINTGSCVLDKGCAPGAPASPLTISVS 656
 Db 590 IPRCPHILDBDGLT-LMKYADKKGGCQCPGHCNCTGCGSPGLSGCGD-IVSHSLAVG 647
 Qy 657 AVNGILLVVLGVVFGIL-----IENNFTVSPMLRVKVSASHLEPITPGAMPNOAQR 711
 Db 648 LVSGLLITVAVALLIIVLLRRRRIRKRRRTICLQEKEL-----VEPLTPGQAPNOAFLE 703
 Qy 712 ILKELTKVYKVLGSCAGFYKGIWIPDGENVAKIPAIKYLENTSPKANKELDEAYV 771
 Db 704 ILKTEFERKDRVLGSCAGFYKGIWIPDGENVAKIPAIKYLENTSPKANKELDEAYV 763
 Qy 772 MAGVSPVSRPLGLCTSTVQVLTQMLPFGCLLDHRENGRGLSGDILNMCQIAKGM 831
 Db 764 MASVDHPVCRLLICITSAVQVLTQMLPFGCLLDHRENGRGLSGDILNMCQIAKGM 823
 Qy 832 SYLEDVLRVHRDLAARVLYKSPHVKITDPGLARLDIDETEHADGKVPKIMMALES 891
 Db 824 NYLEERHLVHRDLAARVLYKSPHVKITDPGLARLDIDETEHADGKVPKIMMALES 883
 Qy 892 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPARIPDLEKGERLPPOPICTIDVYM 951
 Db 884 ILQWTFHQSDVMSYGVTVWELMTFGAKPYDGIIPARIPDLEKGERLPPOPICTIDVYM 943
 Qy 952 IMVCKWMTDSRCRFRFELVSEFSRMAPDQRFVIONEDLGAPSPIDSTFFYSLEDD 1011
 Db 944 IILCKWMTDPSRRFRFELVSEFSRMAPDQRFVIONEDLGAPSPIDSTFFYSLEDD 1000
 Qy 1012 MGDVLDAEYLVPOQGFPCDPAPAGAGMVRHRRSSSTRSGGDLTLGLEPSEEARPS 1071
 Db 1001 -DYVDADEVLLPFYKRI-----NRQGS-----E 1021
 Qy 1072 PLASEGAGSDVPPDGLCMGAAGLQSLPTHDSPLOKYSDDPVP-PLPSTQCYVAPLT 1130
 Db 1022 PCIPPTGH-----PVRENSITTRNISDPQNALEKDLQSH----- 1056
 Qy 1131 CSPOPEYVNPQDVAPQP-----PSPRE-----GPUP-AARPAAGATLERAKTLPSPKNG 1177
 Db 1057 -----EYVNPQGESSTSLSDIYNPNVEDLTDGMGPVSLSSQAEITNFSREYLTNQN 1111
 Qy 1178 VKQVFAFGAVENDEYLTPOGGAAPOHPHPAPASPAFDNLTYWDQDPPEKAPSPFFKG 1237
 Db 1112 L-PLVSSGSDMDPDY---QAG-----YQAP-----LPQTGALTGNGMF 1146
 Qy 1238 TPTAENPEYLG 1248
 Db 1147 LPAENLEYL 1157

RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90311312; PubMed=2164210;
 RA Tloman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
 RA Todaro G.J., Shoyab M.;
 RT "Molecular cloning and expression of an additional epidermal growth
 RT factor receptor-related gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
 RN [3]
 RP SEQUENCE FROM N.A. (SHORT FORM).
 RC TISSUE=Placenta;
 RX MEDLINE=93282822; PubMed=7685162;
 RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
 RT "c-erb3 gene encodes secreted as well as transmembrane receptor
 RT tyrosine kinase";
 RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
 CC SECRETED (SHORT FORM).
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
 CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P65 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
 CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M29366; AAA55790.1; -
 DR EMBL; M34309; AAA55979.1; -
 DR EMBL; S61953; AAB26935.1; -
 DR PIR; A36223; A36223.
 DR HSSP; P11362; 1FGK.
 DR Genew; HGNC:3431; ERBB3.
 DR MIM; 190151; -
 DR InterPro; IPR000494; EGF_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TYK; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; FALSE NEG.
 DR PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Alternative splicing.
 KM SIGNAL 1 19
 FT CHAIN 20 1342 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 644 664 POTENTIAL.
 FT DOMAIN 665 1342 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 709 966 PROTEIN KINASE.
 FT NP_BIND 715 723 ATP (BY SIMILARITY).


```

FT BINDING 742 742 ATP (BY SIMILARITY).
FT ACT SITE 834 834
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 214 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 ELISGVYIEKNDKLCMDTIDMRDIDRADIIVVNDGR
SC -> GQFVPSGGLTPOADOMVLLDDPFRLLTSLASAK
VPTLAV (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

Query Match 35.1%; Score 2389.5; DB 1; Length 1342;
Best Local Similarity 40.0%; Pred. No. 2.2e-121;
Matches 527; Conservative 195; Mismatches 453; Indels 141; Gaps 35;

QY 10 GLLALLPGAA--STVCTGTGDKLRLPASPTHLMLRLHYOGCQVVOGQNLLEYLPT 67
DB 11 GLLFSLARGESEVNSQAVCEGTUNGSLVTGDAENQYOTLYKLYERCEVWNGNLEIVLTGH 70
QY 68 NALISFLQDIQEVGVYLIANOVQVRLQRLTVRGTOLEFEDNYALVALDNDPPLNNTT 127
DB 71 NADISFLQDIQEVGVYLIANOVQVRLQRLTVRGTOLEFEDNYALVALDNDPPLNNTT 125
QY 128 PVTGASPGRLRELRLRLLEILKGLVLIQNPOLCYOTILMNDIFKNNQALTLIDTN 187
DB 126 ----NSSHALRQRLTQLTLEILSSGVYIEKNDKLCMDTIDMRDIYDRD--AEIVVKD 178
QY 188 RSPACHPCSMCKSRGWCSESSDQSLRTVCAGG--ARCKGRLPDCHEQCAAGCTG 246
DB 179 NGRSCPCHEVCCK--KQMGPSSEDCQTLTKITCAPQNGHCFCGPNPQOCHHDCAGGCSG 237
QY 247 PKISDCLACTLHFNHSGICEHCPALVTYNTDTEESMBNPEGRTYFGASCYTAAPYVLTST 306
DB 238 PQDTCDFACHHFNDSGACVPRCPQLVYNNKTLFLBENPHTKYQGVGVCAVSPHNHV-V 296
QY 307 DVSGCTVCPPLHNOEYLAEDGTQCEKCSKPCARVCCGLMQYIKANSK--IGITELE- 363
DB 297 DQGSQVACPDKMEVD--KQGLKMKCECGGLCPACAGTG----SSSRQOTVDSNDIG 350
QY 364 FAGCKITFGSLAFLPESFGDPPASNTAPLQEQHVEVTELEITGYIYIAMPDSLPLDS 423
DB 351 FVNCSTKLGNLDFILTGNQDPFKIKIPALDPEKLNVRTRTRETITGYINTISGWSPPHNFS 410

```

```

QY 424 VFQNTQVIRGRILHNGAVS-LTLQGLISMLGLSLRLSRLSGLALIHNTLCEVHTVPM 482
DB 411 VFSNLTITGGSLVNRGFSLLMKUNLNTSLQFRRLKELISGRITISANROLCHYHSLNW 470
QY 483 DQLEFNPHQALHTA--NRPEDEVGEGLACHOLCARHCWKGPRTOCVNCSQFLRGQEV 541
DB 471 TVLGRPTREERDITKHNRRPRDCVAEGRCVPLCSSGCGMGPQOCLSCRNVSGGVV 530
QY 542 EECRLQGLPREYVNAARCLPCHPECPQNSVTCFGEADQCVACAHYKDPFVARGP 601
DB 531 THCNFLNGEPREFAHEABCFSCHPCEQMEGATNGSGSDTCAQACAFRDPCHVSCSP 590
QY 602 SGVCKPDLSPMPKPEDEGACOPCPINCTHSCVDLDDKCGPABORA----SPLTSIVA 657
DB 591 HGVLG--AKGPYKXPDVQNECRPCHEKCTQCKPRLQDCLQGLVILGKHTLMALTV 648
QY 658 VVGILLV-VLGVV----GILFNPNFVSFRLVRPKXSASLLEPLTSGAMPNOQMNI 712
DB 649 IAGLVIFEMVLQGTFLVWRGRRIQKRAMRRYLE---RGESIEPLDPS-EKANKVLARI 703
QY 713 LKETELRKVVLGSGAFGVYKGIWPDGENYKIPVAIKVIRENTSPRANKELDEAYVM 772
DB 704 FKETELRKVLGSGVFGVHKGVMIPEGSEIKIVCIKVIDKSGRQSFQAVTHMLAI 763
QY 773 AGVSPVYSRLIGLCTSTVOLVQLMFPGCLLDHVENRGRFLSGQDLNMCQIAKMS 832
DB 764 GSLDHAHIVRLIGLPGSSSLQVLTQYLPGLSLDHHVORHGAIGQLLMGVOIAKGVY 823
QY 833 YLEDRIYVRDLAANNVLYKSPNHKINDPGARLLDDETEYHADGGVPRKMMALESI 892
DB 824 YLEBGMVHRNLAAKNVLLKSPQYQVADPFGADLLPDDKOLLYSEAKTPIKMMALESI 883
QY 893 LRRPFTHOSDVNSYGVYTWELMTFGAKPYDGIIPAREIDPLEKGERLPORPICTIDVMI 952
DB 884 HNGKTHQSDVMSYGVYTWELMTFGAEPYAGRLAEVPRDLEKGRLAQPOCTIDVYV 943
QY 953 MYKCMWIDSECCPRRELVSFSSMAARDPQRRVYVQNEDLGA--SPLDSTFYSLLED 1009
DB 944 MYKCMWIDENIRPTEKELANETFRMARPPRLVLIKRES--GGIAPGPRPHGLTKKLEE 1002
QY 1010 DDMGDLVDAEYLVQCGFPCDPAPAGAGWNNHRRSSSTSGSGDILLGLEP--SEEA 1068
DB 1003 VLEPELDLDDLEAED-----NATLTLSALSPLVGTINRRPG 1043
QY 1069 PRSPPLAPSGAGSDVFDGLGKAAGLQSLPTH--PSFLQRYSEDPTVPLP-----SE 1121
DB 1044 SOLLSPPSSGY--MPMNGNLGSCQBSAVSGSSERCPRVSLH-----PMRGCIASES 1096
QY 1122 TDGYVA-----PLCSPPQE-----YVNPQVRRPQPPSRREGP----- 1154
DB 1097 SEGHTVGSFAELQEKVSMKRSRSPRPBGDSAYHSORHSLTLPTVPLSPFGLEEDV 1156
QY 1155 ----LPAARPAQATLERAKTLP--GKNGVY-----KVFARGAVERNPEYVLPQGAAP 1203
DB 1157 NGVYMPDTHLKGTSPSREGTLLSSVGLSVLGTDEED-----EEVEYNNRRRHSP 1208
QY 1204 QHPPRPAPSPAFDNLVYND-----QDPREKGRAPSTFGTIPAKENPEYL 1247
DB 1209 -PHPPRPSLEELGYEYMDVSGDLSASLSTQSCPLHVPVIMPTAGTTPDEDEYEM 1263

RESULT 10
ERR3 RAT
ID ERR3 RAT STANDARD; PRT; 1339 AA.
AC 062759; 062955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (C-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; Tissue=Liver;
 RC MEDLINE=96086535; PubMed=8522190;
 RA Heliyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Kojand J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RT recombinant protein.";
 RL Gene 165:279-284(1995).
 RN [2]
 RP REVISIONS TO 85; 513 And 565.
 RA Heliyer N.J., Kojand J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; Tissue=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroligin and their putative receptors, ErbB2 and
 RT ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTKA.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 210 218 BY SIMILARITY.
 FT DISULFID 214 226 BY SIMILARITY.
 FT DISULFID 227 235 BY SIMILARITY.
 FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1028 1028 L->P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402B5FD1E CRC64;
 Query Match 34.0%; Score 2315.5; DB 1; Length 1339;
 Best Local Similarity 40.4%; Pred. No. 2.1e-117;
 Matches 550; Conservative 171; Mismatches 431; Indels 165; Gaps 36;
 QY 3 LALCRWGLLLALPPGAA---STQVCTGTDMKRLRPAPEHLDMLHLYGCGQVQVGN 59
 DB 7 LQVLC---FLSLARGSMGNSQAVCPRTGLNGSLVTDADQVQVLYKLYEKEVWGN 62
 QY 60 LELTLPPTVASLFIQDIOEVGYVLIANQVRQVPLQRLRVRGTQLEFNVLAVALDN 119
 DB 63 LELVLTGHADSLFIQWIREVGYVLIANMERSVPLRLRVRGTQVYDQKFAIFVM-- 120
 QY 120 GDP LNNTPTVTGASPEGREQLRLRLTEILKGVLIQRNPQLCYDPTLMDKIFPKNNOL 179
 DB 121 ---LNVNT---NSSHALRQLKFTQLTETLSGVYIEKXDKLCHMDTIDWRDIVVR-- 170
 QY 180 ALLTIDNRSRACHPQSPMCKGRMGSESSDQSLTRTVACGGC-ARCKGPLPTDCHE 238
 DB 171 GAELIVKNGANGCPCHVEYCKG-RCKGPPDDCQILTKTICAPQCNRCFGFNPQCCHD 229
 QY 239 QCAAGCTGPKSHDCLALHFNHSGICEHCPALVTYNTDFESMNPREGRYFGASCYTA 298
 DB 230 ECAGGCGSPQDTPDFCAFRFRFNDGACVPCPEPLVYNNKTLFQLEPNPHTKYQYGGVCVAS 289
 QY 299 CPNNTYSTVGSCTLVCPAHNOEYVAEDSTQCEKSCKPCACVAVGL--GMQYIANKSKF 356
 DB 290 CPENFV-VQOTCVRACPDKMEVD-KHGLKMCCEPGGLCPACGCTGSGASYQVDSN 347
 QY 357 IGTLEFPAQCKKIFGSLAFLEPSFGDPASNTAPLOEQLOVFTELEITGYLYISAWP 416
 DB 348 ID----GFVNCCKIIGLNDLFLTGLVNDPMHKI PALDPEKLVNFTVREITGYLNIQSWP 403
 QY 417 DSLPDLVSQONQVIRGRILNHGANS-LTLOGLGISMVLGRLRELGLSLAIHNTHLC 475
 DB 404 PHHNSVFSNLTITGGRSLYNGFSLLMKNLNVTSGFRSLKELISAGRYVTSANQOQC 463
 QY 476 FVATVPMQDLFNNPQALHTA-NRPEDECVGEGLACHOLCARCGCMGPGPTQVCNCSQF 534
 DB 464 YHSLNMTTLTLGSPSERIDIKYDRPLSGCLAEKGVCPDLSSGGCMGPGQCLSCNRY 523
 QY 535 LRGQECVEBCRYLQGLPREYVNAHCLCPHECQFQNGSVTCFGEADQCAVACAHYKDP 594

```

Db      | 524 SRGVCVTHNFIQGEREFVHAQCFSCHEPCLPNEGSTGCGSSDACARCAHFRDGP 583
Qy      | 595 FCVACRPSGVKPLSTWPIKFPDEGACOPCPINCTHSC--VDLDDKCPAQRASPLT 652
Db      | 584 HCVNSCPHGILG--AKGPIKYPDAQNECRPCHECTOCGNGEPLDDCQGAELVSKPH 641
Qy      | 653 SIYSAVVGI-LLVVIGVVF---GILFNNFTVSPMLRVPKYSASHLEPLTPSGAMPNQ 707
Db      | 642 LVIAVYGLAVIIMILGSGFLYWRGRIRQKRAMRYLE---RGSISFLDPS--EKAKK 696
Qy      | 708 AQMRILKETELRKVKVLSGAFGTVYKGIWIPDENVKIPVAIKVIRENTSPKANEILD 767
Db      | 697 VLARIFKETELRKVKVLSGAFGTVYKGIWIPDENVKIPVAIKVIRENTSPKANEILD 756
Qy      | 768 EAVVMAGVSPVYSLIGICTSTVCLVLTQMLVYGGCLDHPVRENRIKLSODLLNCKOI 827
Db      | 757 HMLAVGSLDHAHIVRLGLCPGSSISLQVLTQYLPGLSLDHYKHRETLGQILLMNGVQI 816
Qy      | 828 AKGMSYLEDVRLVHRDLAARNVLVKSPNHVKTDFGLARLLDIDETRYADGQKVPDKM 887
Db      | 817 AKGMYLLEHSMVHRDLARNWMLKSPSQVQVADDFVADLLPDDKQLLHSEAKTPIKMM 876
Qy      | 888 ALSEILRRFTHOSDWSYGVTWELMTGAKPYDGPAPETIDLEKGRLEPQPICTI 947
Db      | 877 ALSEIHFGKTHOSDWSYGVTWELMTGAEPLAGRLAEIPDLLEKGRLEPQPICTI 936
Qy      | 948 DVTMIMVCKMMIDSECPREPRELVEFSRMAPDPORFVVIQNEIDLGASPLDSTFRRSL 1007
Db      | 937 DVTMIMVCKMMIDSECPREPRELVEFSRMAPDPORFVVIQNEIDLGASPLDSTFRRSL 993
Qy      | 1008 EDDMDGDVDAEYLVPQGFPCPDPAFGAGVHHRHRSSTRSGGDLTCLGSESEF 1066
Db      | 994 TTREL-----QEALEPEL-----DLDLLEAEER 1019
Qy      | 1067 -----EAPSPPLAPSEG-----AGSDVFDGLGMAKGLQSL 1099
Db      | 1020 LATSLSGALSPTGTLTRPGSCSLSPSSGVMFMNOSLGEACIDSAVLGEGFSRPE 1079
Qy      | 1100 PTHDPSPLQRYSEDPYLPSEFDGVY---ADP-----TC-----SPQF-----YVN 1139
Db      | 1080 SLH-FIPRGR-----PASESEGHVYTGSEALOEKVSVCRSRSRSPRPGSDAYS 1131
Qy      | 1140 QPVRPQPPSPRPGP-----LPARPAQATLBRXKTLSP-GKXGVY-----CD 1181
Db      | 1132 QRSLSLTPVTPLEPPLGEEEDGNGYMPDTHLGASSRECTLSSVLSVLTGEDEED 1191
Qy      | 1182 VFAFGAVENPEYLTTPQGAAPQPHPP 1208
Db      | 1192 -----EVEYVNRKRGRSP-PRPP 1209

```

```

RESULT 11
EGFR DROME STANDARD; PRT; 1426 AA.
AC P04412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EG 2.7.1.112) (Egfr)
GN (Gurkin receptor) (Tropo receptor protein) (Drosophila relative of EBBB).
OS EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in

```

```

RT subdomains of the receptor protein."
RL Genetics 137:531-550(1994).
RN [2]
RP REVISIONS.
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Lively E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains."
RL Cell 40:599-607(1985).
RN [4]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; Tissue=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schjeter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts."
RL Cell 46:1091-1101(1986).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila."
RL Dev. Biol. 205:129-144(1999).
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe O., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrita J.F., Agbayani A., An H.-J., Andrews-Pfannkuh C., Baldwin D.,
RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benoe P.V., Berman B.P., Bhargava D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Feyraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mikhlin N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pachet J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spierckas R., Tecor C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Williams S.M., Woodage T., Worley K.C., Xu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [7]
RP SEQUENCE OF 959-1078 FROM N.A.

```

CC STRAIN=Daekyangeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S., III, Bildeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
RT growth factor receptor.";
RL Nature 314:178-180(1985).
RN [8]
RP SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RP ANALYSIS.
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "Interallelic complementation among DER/Flb alleles: implications for
RT the mechanism of signal transduction by receptor-tyrosine kinases.";
RL Genetics 129:191-201(1991).
RN [9]
RP REVIEW.
RX MEDLINE=97248481; PubMed=9094709;
RA Pettimon N., Perkins L.A.;
RT "There must be 50 ways to rule the signal: the case of the Drosophila
RT EGF receptor.";
RL Cell 69:13-16(1997).
CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GRKRN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSROA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGE OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on the
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-stb.ch/announce/>)
CC or send an email to license@isb-stb.ch.
CC -----
DR EMBL, AF052754; AAC08536.1; -
DR EMBL, AF052753; AAC08536.1; JOINED.
DR EMBL, AF052754; AAC08535.1; -
DR EMBL, AF052752; AAC08535.1; JOINED.
DR EMBL, K03054; AAA51462.1; -
DR EMBL, K03417; AAA51460.1; -
DR EMBL, K03416; AAA50965.1; -
DR EMBL, K03418; AAA51461.1; -
DR EMBL, AF109077; AAD26134.1; -
DR EMBL, AF109078; AAD26132.1; -
DR EMBL, AF109082; AAD26133.1; JOINED.
DR EMBL, AF109078; AAD26133.1; -
DR EMBL, AF109084; AAD26133.1; JOINED.
DR EMBL, AF109079; AAD26130.1; -
DR EMBL, AF109081; AAD26130.1; JOINED.
DR EMBL, AF108079; AAD26131.1; -
DR EMBL, AF109083; AAD26131.1; JOINED.
DR EMBL, AF109080; AAD26135.1; -

Query Match	Best Local Similarity	Score 1924;	DB 1;	Length 1426;
Matches 458;	Conservative 189;	Mismatches 433;	Indels 348;	Gaps 41;
24	QVCTGTDMKLLPLPSPETHLMLRHLVYOGCGVQGNLELYLPT-NASLSFLQDIQEVQ	82		
100	KICIGTKSRLLSVPSKSHENHYNLRDRYNTCTGYVQGNLKLTYLMPENLIDLSFLDNIREVTG	159		
83	YVLIANOVROYPLRLRIYVGTQLF-----EDNYALAVLNDGPRINNTPTVTSRPGSL	137		
160	YILSHVAVKVKVPEKQIIRGRTLSLSVEEKALV-----TYSRM	203		
138	RELQSRSLTEILKGVGLIQRNPOLCYQDTIEMKIDFHKNQALTLIDITNSRACHPSR	197		
204	YTLFPLDLVDVLNGVGFHNNYNLCHMTIQMSLVSNGTDAYVYDTPARERCPKCHE	263		
198	MCKGSRGWGESSEDCSLTRIVCAGCA-RCKGFLPDDCHBQCAAGTGRKSHDCLAC	255		
264	SCTHS-CWGEGRKNQKFSKLTCSQCAGRGYGRKPCCHLFCAGGCTGPTQKDCIAC	322		
256	LHNHNSGICELHCPALVTYNDTPEMSNPEGRYFGASCYACAYNLSTVDSCTIWC	315		
323	KNFPEDAVSKECPMKRKNPTTYLLENPEKAYVGTAYKECP-GHLRDNQACVASC	381		
316	PLHNOVTAEDGTQRCXKSKRCARVCYGLGMQYIKANSKFIGITEL-----EPAGCK	369		
382	PQDKMDKGE-----CVPNGCPKTC-----PGVTLVLAGNIDSPRNCV	422		
370	IFGSLAPFBSADG---DPAANTA-----PLQPELOVFEPLLEITGLYLSAWPDSLDP	421		
423	IDGNRIILDQTSFGQDVYANTMGPRYIPLDPERREVSIVKELTGIANLGGTIPQRN	482		
422	LSVFONLQVIRGIRLIHNGAY-SLTQGLGISWLGSRSLRELSGGLALIHNNHLCFYHTV	480		
483	LSYFNLELTHIRQLMESWPAALAVKSVLSLEARNIKQISGSGVVIQHNBDLCYVSN	542		
481	PPWDQLFRNPQHALLTTPANRBEVCSQEGLAGQGLCARHGCHWGRGTQCVNCSQFLRGQEC	540		

```

Db      543 RWAIAIKPEQKQWVWVNNRADICENKNGTICSDQCNEDGQWGGTQCLTCNKNFNNGTC 602
Qy      541 VEECRVLQGLPREYVNAARHCLPHPEQCPONGSVTGPBPADOCVACAHYKDPPEVAAAC 600
Db      603 IADCGVISAAYK--FDNRCKIKHPECR-----TCNGAGADHCQECVHRDQHCVSBC 654
Qy      601 P-----SGVK-----PDL 608
Db      655 PKKKYNDRGVCRECHATCDCTGPKDTIGACTTCNALINNDATVKKCLKDDKCPD- 713
Qy      609 SYMPIWKE--PDEEGACOP-----CPI-----NCTH----- 632
Db      714 GY--FMEYVHPQEGSLKPLAGRAVCRKCHPLCELCTNYGHEQVCSKTHYKREQCET 771
Qy      633 -----SC-----VLDLKG----- 641
Db      772 ECPADHYTDEEQECCFQRPHECNCCTGPGADCKSCNFKLFPANETGPNVSTMCNCTG 831
Qy      642 -CEAQR-----ASPLTSIVSAVGI-LLVVLGVVFGILLFNNFTVSEW 684
Db      832 KCELEMHVNVQYTAIGPYCAASPBRSSKITANLDVNMFIITGAVLVPTICILCVTYI 891
Qy      685 LRPKVSASHL-----EPLTSGAIPNOQMILKETLKRKKTGSAFET 731
Db      892 CRQKQAKKETVKTMTALSCEDSEPLRPSNIGANLCKLIVDAELRKGVGWGAFG 951
Qy      732 VYKGIWIPDGENYKIPVAIKVLRBNTSPKANKEILDEAYYVAGVSPVYSRLIGLCTST 791
Db      952 VYKGVWVPEGENYKIPVAIKELKSTGAESEFLEAYIMASEEHVNLKLLAYCMSQ 1011
Qy      792 VQVLQMPYGCILLDHYRENRKGLSGODLLNMCQIAKMSYLEDVLYHRDLAARNVLY 851
Db      1012 IMMLITQMPGLCLLDYVRNNRDKIGSKALNMSTQIAKMSYEEKRLVHRDLAARNVLY 1071
Qy      852 KSPNHYKITDFGLARLDIDETEVHADGKVPIKMMALLESILRRFTHOSDVSYGVTVW 911
Db      1072 QTSLSYKIDTDFGLAKLSSDSNRYKAAAGKMPKMLALBEIRNRVFTSKSDVAFGVITW 1131
Qy      912 ELMTFGAKPYDGIIPAREIPPLLEKGERLPORPCTIDVYIMWKCMIDSECRPRELEV 971
Db      1132 ELTTFGQRPHEINIPAKIDPLIEVGLKEBPETCSIDYCTLLSCMHLDAAMPRTKQULT 1191
Qy      972 SEFSMARDPQRFVIVONEIDLG--PASPLDSTFYRSLIEDD--DWGDLVDAEYLVPOQ 1026
Db      1192 TVFAEPRARDGRITLALPGDKFTLPA-----YTSODEKDLIRKLAPTTDGEALAKPD 1244
Qy      1027 GFECPPDPAFGAGGVHRRHRSSTRSQGGDLTLGLEPSEEAR-----RSLAPSEGAG 1080
Db      1245 DYLOPKAAPGPS-----HRTDCT-----DEMFKLNRVCMDPSNKNKSTG 1283
Qy      1081 SDVPFG--DLGMAKAGLQSLPTHDSPLQRYSEDPVPLPSETGQVAPLTCSPQPEY 1137
Db      1284 DDERDSAREVGVGNLR-----LDLPVEDDYIMP--TCQGGFPNN 1321
Qy      1138 VNQDPVPPCPSPREGEPLPAARPAAGATLEBAKTLSPGKGVNVDPVAFAGGAVENPEYL- 1195
Db      1322 NNNNN-----NNQNNMAAVGAAGYM-----DLGVPVSDVNPPEYLLN 1360
Qy      1196 --TPQGAAPQPH-----PPAPSP-AFDNLYTWD 1222
Db      1361 AQTLLGVGESEIPQTIGIPVMGGPGTMEYKVPMPGSEPTSSDHEYND 1408

```

```

RESULT 12
ERBB ALV
AC      21-JUL-1986 (Rel. 01, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DE      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
V-ERBB.

```

```

OS      Avian leukosis virus.
OC      Viruses: Retroid viruses; Retroviridae; Alpharetrovirus.
OX      NCBI_Taxid=11864;
RN      (1)
RX      SEQUENCE FROM N.A.
RP      MEDLINE=85228222; PubMed=2988784;
RA      Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA      Crittenden L.B., Raines M.A., Kung H.-J.;
RT      "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT      processing and promoter insertion result in expression of an
RT      amino-terminated EGF receptor."
RL      Cell 41:719-726 (1985).
CC      -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC      IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC      AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC      THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC      c-ERBB GENE, WHICH LEADS TO c-ERBB ACTIVATION.
CC      -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC      PROTEIN.
CC      -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb.ch).
CC
CC      EMBL: M10066; AAA8763.1; ALT_INIT.
CC      PIR: A00643; TVCHLV.
CC      PIR: B00643; TVFVLV.
CC      HGSP: P11362; 1FGK.
CC      InterPro: IPR000719; Euk_Pkinase.
CC      InterPro: IPR001245; Tyr_Pkinase.
CC      Pfam: PF00069; Pkinase; 1.
CC      PRINTS: PR00109; TYRKINASE.
CC      ProDom: PD000001; Euk_Pkinase; 1.
CC      SMART: SM00219; Tyrc; 1.
CC      PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC      PROSITE: PS00109; PROTEIN KINASE TYR; 1.
CC      PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC      Transfaser: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      KM Glycoprotein; Phosphorylation.
CC      FT DOMAIN 132 399
CC      FT NP BIND 138 146 ATP (BY SIMILARITY).
CC      FT BINDING 165 165 ATP (BY SIMILARITY).
CC      FT ACT SITE 257 257 BY SIMILARITY.
CC      FT ACT SITE 257 257 BY SIMILARITY.
CC      SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

```

```

Query Match      24.5%; Score 1672.5; DB 1; Length 634;
Best Local Similarity 50.8%; Pred. No. 3.9e-83;
Matches 359; Conservative 77; Mismatches 147; Indels 123; Gaps 17;

```

```

Qy      587 CAHYDPPFCVARGCSGYKPLDSYMIKFPDEBEGACOPPTNCHSVDDDKCPAPQ 646
Db      3 CAHFDGPRCVACAGVLENDTL-VKVIADANAACVQLCPNCTRGCKGGLBCP-- 58
Qy      647 RASPLTSIVSAV--GILLVVLGVVFGILLFNNFTV--SFWLRVPKVSASHLEPLPSGAM 704
Db      59 NGSKTPTSAAGVGGLLCLVYVVGIGILYLRHRHVRKRTLRLIQERELVEPLPSSGEA 118
Qy      705 PNOQAMRIKETELKAYVVGSGAGTGYKGIWIPDGNVYKIPVAIKVLRBNTSPKANKE 764
Db      119 PNOAHLRIKETEFKVVVLGSGAGTVYKGIWIPDGNVYKIPVAIKVLRBNTSPKANKE 178
Qy      765 ILDEAYYVAGVSPVYSRLIGLICITSTVQVLTQMLPYGCLLDHYRENRKGLSGODLLNMC 824
Db      179 ILDEAYYVAGVSDNPRVCHLLIGLICITSTVQVLTQMLPYGCLLDYIREHNDNIGSYLLNMC 238
Qy      825 MOIAKMSYLEDVLYHRDLAARNVLYKSPNHVKITDFGLARLDIDETEVHADGKVP 884

```

```

Db      239 VOIAKMNYLEERLVRDLAARNVVKTPQHVKTIDFGLAKLGLADEKEVHAEGSKVPI 298
Qy      885 KMALESILRRRFTHSDVWSYGVYWEMLTFGAKPYDGIIPAREIDLLEKGERLPQPI 944
Db      299 KMALESILRRITTHSDVWSYGVYWEMLTFGSKPDGIPASEISSVLEKGERLPQPI 358
Qy      945 CTIDVYMIWKCWMIIDSECRPRRELVSSESMARPPQRFVYIQ-NEDLGASPLDSTFY 1003
Db      359 CTIDVYMIWKCWMIIDSECRPRRELIAEFSKMARPPRYLVIQGERHMLPSPTSKFY 418
Qy      1004 RSLLEDDMGDLVDAEYLVLPQGFCCPDPAAGGVHHRHSSSTRSGGDLTLGLEP 1063
Db      419 RTLMEEDMEDIDVADAEYLVPHQGF-----NSPST----- 449
Qy      1064 SEEARPSPL-----APSEGAGSDVDGDLGMAKAGLSLPHDPSPLQRYSEDTVPL 1118
Db      450 -----SRPLSLSLSTSNNSATNCID-----RNGQGHPRREDSFQRYSSDPTGNF 496
Qy      1119 PSET--DGYYAPLTCSPQPEYVNPQDVPRQPSPREGLPAPAPAGATLERAKTSPGKN 1176
Db      497 LEESIDDGFL-----PAREYVNO--LMPKKPS-----TAMNOC 527
Qy      1177 GYVKDVF-----AFGAIVENPEYLTPOGGAAPQHPPEAFSPAFDNLVYM 1221
Db      528 QIYNNISILTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPYW 579
Qy      1222 DQ-----DPE-----RCAPSTFKGTPTAENPEYLGDPV 1252
Db      580 IQSGNHQINDNPDYQODFLPNETKPNGLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AYIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain Es4).
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
   gene family."
RL Cell 35:71-76(1983).
RN RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Dejure B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
   new type of oncogene."
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
   tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
   ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
   IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
   RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYU.
DR HSSP; P11362; IEGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD067450609 CRC64;

Query Match 23.9%; Score 1626; DB 1; Length 604;
Best Local Similarity 50.7%; Pred. No. 1,2e-80;
Matches 349; Conservative 74; Mismatches 140; Indels 126; Gaps 16;

Qy 587 CAHYKDPFCVARGCSGVKPDLSYPMKFPDEBAGACQPCPINCCHSCYVDLDDKCPAQ 646
Db 3 CAHFIDGPHCVACAGVAGENDTL-VKRYADANAVCOLCHNCRTGKGPGLECCP--- 58
Qy 647 RASPLSTYSAAV-GILLVWLVGVVFGILFNNFTV-SFWLRVPKVASHLEPLTPSGAM 704
Db 59 NSKTPSLAGVAGVGLCLVVGIGLYLRNRHIVKRTLRLLQGERLVEPLTPSGA 118
Qy 705 PNOAQMRILKETELRKVYKLGSGAFGYKGIWIDGENVKIPAIKVLRENTSPRANKE 764
Db 119 PQOHLRIKTEFEKKVKYLGSGAFGTIKGIMDEGEVKIPAIKELRENTSPRANKE 178
Qy 765 ILDEAVMAGVSPYSLIGLCTSTVQLVQLMPYGLLDHVENRGLSODPLNLC 824
Db 179 ILDEAVMASVNPFCRLGICLSTVQLVQLMPYGLLDHVENRGLSODPLNLC 238
Qy 825 MQIAKMSYLEDRVLRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGKVP 884
Db 239 VOIAKMNYLEERLVRDLAARNVVKTPQHVKTIDFGLAKLGLADEKEVHAEGSKVPI 298
Qy 885 KMALESILRRRFTHSDVWSYGVYWEMLTFGAKPYDGIIPAREIDLLEKGERLPQPI 944
Db 299 KMALESILRRITTHSDVWSYGVYWEMLTFGSKPDGIPASEISSVLEKGERLPQPI 358
Qy 945 CTIDVYMIWKCWMIIDSECRPRRELVSSESMARPPQRFVYIQ-NEDLGASPLDSTFY 1003
Db 359 CTIDVYMIWKCWMIIDSECRPRRELIAEFSKMARPPRYLVIQGERHMLPSPTSKFY 418
Qy 1004 RSLLEDDMGDLVDAEYLVLPQGFCCPDPAAGGVHHRHSSSTRSGGDLTLGLEP 1063
Db 419 RTLMEEDMEDIDVADAEYLVPHQGF-----NSPST----- 449
Qy 1064 SEEARPSPL-----APSEGAGSDVDGDLGMAKAGLSLPHDPSPLQRYSEDTVPL 1118
Db 450 -----SRPLSLSLSTSNNSATNCID-----RNGQGHPRREDSFQRYSSDPTGNF 496
Qy 1119 PSET--DGYYAPLTCSPQPEYVNPQDVPRQPSPREGLPAPAPAGATLERAKTSPGKN 1176
Db 497 LEESIDDGFL-----PAREYVNO--LMPKKSTAM----- 524
Qy 1177 GYVKDVF-----AGAVENPEYLTPOGGAAPQHPPEAFSPAFDNLVYM 1221

```

```

Db      525 -VQNGYNFLISLAIKSLPMDSRKQNSHSAVDNPEYL-----NTNQSLAKTVES 575
QY      1218 LYYWQDPPREGAPSTFKGTPTAENPEY 1246
Db      576 SPYWIQSGNHQ-----INLDNDY 594

RESULT 14
ERBB_AVEIU STANDARD; PRT; 540 AA.
ID ERBB_AVEIU
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN v-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP MEDLINE=87064456; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus--transformed erythroid
RT cells."
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -|- tyrosine phosphate.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31179; AAA42401.1; -
DR PIR: A25231; TVFEH.
DR HSSP: P1362; 1FGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TYKc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE: PS50011; PROTEIN KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NE_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297A0A06B85D CRC64;

Query Match 22.8%; Score 1553; DB 1; Length 540;
Best Local Similarity 53.4%; Pred. No. 8; 8e-77;
Matches 330; Conservativity 67; Mismatches 131; Indels 90; Gaps 14;

QY 587 CAHYKDPFCVACPSGVKXKDLSTYPMKPPDEGACOPCPINCTHSQVDLDDKGPAPG 646
Db 3 CAHFDGPHCVKACPSGVKXKDLSTYPMKPPDEGACOPCPINCTHSQVDLDDKGPAPG 58
QY 647 RASPLTSIVSAVV-GILLVVIGVFGIILFNFTV-SFWLRVQVKSASHLEPLTSGAM 704
Db 59 NSGKTPSIAGVGVGILCLVAVGLIGLYLRHRHIVKRTLRLLQERLEVEPLTSGBA 118
QY 705 PNAQWRILKETELRKVKVLGSGAFGVYKGMIPGNGVNVKPVAKVLRNTSPKANKE 764

```

```

Db      119 PNAQWRILKETELRKVKVLGSGAFGVYKGLMIPEGEKVTIIPVAIKLEAATSPKANKE 178
QY      765 ILDEAYVMAVGSGPYVSRILGICITSTYQVLTQMLPFGCLLDHRENRGRIGSQDLLNWC 824
Db      179 ILDEAYVMAVDNPNVQCLLIGICITSTYQVLTQMLPFGCLLDYIREHNDNGSQVLLNWC 238
QY      825 MQIAKMSYLEADVLRVHDLARNVLVKSPNNHVTITPGLARLLIDDETEHADGKVPY 884
Db      239 VQIAKMYLLEERHMYHRLDARNVLVKTPGHVITITPGLAKOLGADKEKHAEGCKYPI 298
QY      885 KMALESILRRRFTHQSDVMSGYVTWELMTFGAKPYDGPAREIPDLEKGERLPQPI 944
Db      299 KMALESILHRIYHQSDVMSGYVTWELMTFGAKPYDGPAREIPDLEKGERLPQPI 358
QY      945 CTIDVYMTVMCMWIDSCRPREFEIVSEFRMRDPPRFVIO-NEDLGPASPLDSTFY 1003
Db      359 CTIDVYMTVMCMWIDSCRPREFEIVSEFRMRDPPRFVIOGDEKMLPSPTDSKFY 418
QY      1004 RSLLEDDMDGLVDAEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEP 1063
Db      419 RTIMEEDMEDIVDADEVLPHQGF-----NSPST----- 449
QY      1064 SEEBAPRSPL-----APEEGAGDVFDGDLGMAKGLQSLPTHDPSPLOKYSDDPTVPL 1118
Db      450 -----SRTPLLSSATSNSATNCIDRNG-----H----- 476
QY      1119 PSETDGYVAPLTCSPQEPYVNVQPPVPPQSPREGRPLPAPAPAGAT-LEAKTLPGRNG 1177
Db      477 PVEDRGLF-----PAPRYVNG--LMPKRPSTANVQVNYVITSLTSLKPIDSRYN- 527
QY      1178 VKQVFAFGAVENPEYL 1195
Db      528 -----SHSTAVDNEPYL 539

RESULT 15
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC EGFR_CHICK
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -|- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC -|- tyrosine phosphate.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M20386; AAA48760.1; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR01245; Tyr_kinase.
 DR Pfam: PF01030; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU_4.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; PARTIAL.
 KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 30
 FT CHAIN 31 >703
 FT DOMAIN 31 654
 FT TRANSMEM 655 667
 FT DOMAIN 668 >703
 FT DISULFID 197 206
 FT DISULFID 201 214
 FT DISULFID 222 230
 FT DISULFID 226 238
 FT DISULFID 239 247
 FT DISULFID 243 255
 FT DISULFID 258 267
 FT DISULFID 271 298
 FT DISULFID 302 314
 FT DISULFID 318 333
 FT DISULFID 336 340
 FT DISULFID 513 522
 FT DISULFID 517 530
 FT DISULFID 533 542
 FT DISULFID 546 562
 FT DISULFID 565 581
 FT DISULFID 589 592
 FT DISULFID 592 601
 FT DISULFID 605 627
 FT DISULFID 630 638
 FT DISULFID 634 646
 FT CARBOHYD 134 134
 FT CARBOHYD 190 190
 FT CARBOHYD 200 200
 FT CARBOHYD 359 359
 FT CARBOHYD 368 368
 FT CARBOHYD 420 420
 FT CARBOHYD 573 573
 FT CARBOHYD 578 578
 FT CARBOHYD 613 613
 FT CARBOHYD 633 633
 FT CARBOHYD 648 648
 FT NON_TER 703
 SQ SEQUENCE 703 AA; 77427 MW; AEF2DE11B735A690 CRC64;

Query Match 22.3%; Score 1518; DB 1; Length 703;
 Best Local Similarity 43.1%; Pred. No. 9.1e-75;
 Matches 305; Conservative 109; Mismatches 263; Indels 30; Gaps 14;

QY 8 RWGLLALLPPGAA-----STVCTGDMKLRLPASPEHLDMLRLHYOGCCVYQGNLE 61
 DB 13 RGAALVLLLVGLVLCGSAVEEKVCGQNNKLTQLGHVEDHFTSLQRRYNNCEVALSNLE 72
 QY 62 LTVLPTNASLSFLQIDQVQGVLIANQVQVPLQSLRIVRGTOLEFDNYALAVLDNGD 121
 DB 73 ITVEHNEDLTFLEKTIQEVAGVLIALNMVDVIPLENLQIRGNVLYDNSPALAVLSNH 132
 QY 122 PLNNTPTVGTASPGRLRLQSLTEILKGVLIQRNPQLCYQDTILMKDTFHKONQLAL 181

DB 133 -MNTQ-----GLRELPMKRLSELINGVXISNNPKLCNMOTVLANDIIDTSRK-PL 182
 QY 162 TLID-TNRSRACHPCSPMKCKSRCKBESSEDCQSLTRVCGGGA-RKSGELPDDCCHEO 239
 DB 183 TVLDASVASSCPKCHPCTEDHCGAECQNCQTLTKVICAQOCSGRGRGVPSDCCCHQ 242
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFFSMRPEGRYTFGASCVTAC 299
 DB 243 CAAGCTGRESBDCACRKFRRDQATCKQCPPLVLYNPTTYGMVDNVBEKYSFGATCVNEC 302
 QY 300 PNYLSTDVSGCTVCPHINOEVTAEDGTQRCCKSPCARVCGLGMOYIKASKPFGI 359
 DB 303 PHNYVVTDHGSCVRSQNTDYEY--EENGVKCKKCDGLCSKYCNIGIGELKGLIS-INA 360
 QY 360 TELF-FACCKTFGSLAFSPESFGDPAASNTAPLOPELOVFEETLEETGYLYISAMPDS 418
 DB 361 TNIDSPKCTKINGVSLPVAFLGDATKTLPLDPKCLDVFRYVKEISGFLLIQAMPDN 420
 QY 419 LPDLVFOQLQVIRGRILHNGAVSLTLOGLSIWGLRSLRSLRSLGSLALIHNTLCEVH 478
 DB 421 ATDLVAFERLEIRGRTRQHQGYSLAVVNLKIQSLRLSKEISDGLAIIMKNKILCYAD 480
 QY 479 TYPMDQLFRNPQALLHTANRPEDECVGEGLAGHQLCARHGMGSPQCVNCSQFLGQ 538
 DB 481 TMNWSLFPATQSQKTKIIONRKNDCDADRHYVCDLCSDVCGMGSPHCFSCRFPSRQK 540
 QY 539 ECVEEGRVYQGLFREYVNAARHCLPCHPECPQNG---SVTCFPEADOCVACAHYKDPFF 595
 DB 541 ECVKQCNITIGEPREFRENDSKLPCHSECLVNSTANVTTCGPGPDHGMKCAHFIDGH 600
 QY 596 CVARCPGKVPDLSYMPYWKFPDEGACQCPINCTHSCVDLDYGCAPAEQASPLTSIV 655
 DB 601 CVKACPAGVLGNDPL-VWKYADANAVQLCHPNTCTRGCKGPGLEGCP---NGSKTPISA 656
 QY 656 SAVV-GILVVVLGVVFGILIFNNFTV-SFMLRVPKVSASHLEPLTP 700
 DB 657 AGVGGLLCLVVGGLGILYLRRIIVKRTLRRLQERELVEPLTP 703

Search completed: July 22, 2003, 09:18:21
 Job time : 20.625 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.75 Seconds
(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-653-675-14
Perfect score: 6814
Sequence: 1 METALCRMGDLALPLPGA.....TFKGTPTAENPEYLGDVAV 1253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL.21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp Vertebrate.*
14: sp Unclassified.*
15: sp_virus.*
16: sp_bacteriopl.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6097	89.5	1259	6 O18735	O18735 canis fam1
2	3060	44.6	1209	11 Q9GX70	Q9GX70 ratcus nov
3	3038	44.6	1210	11 Q9EP98	Q9EP98 mus musculu
4	2647	38.8	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2636.5	38.7	1137	13 Q9W6F6	Q9W6F6 gallus galli
6	2265	33.2	1338	13 P79754	P79754 fugu rubrip
7	2012.5	29.5	1433	5 Q9BIF9	Q9BIF9 anopheles g
8	1871	27.5	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.5	357	11 Q8R2X1	Q8R2X1 mus musculu
10	1697.5	24.9	412	4 Q8WYV0	Q8WYV0 homo sapien
11	1643	24.1	729	15 Q86712	Q86712 avian rous-
12	1641	24.1	567	15 Q86714	Q86714 avian rous-
13	1576.5	23.0	962	15 Q64895	Q64895 avian eryth
14	1568	23.0	545	15 Q85468	Q85468 avian eryth
15	1506.5	22.1	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1490.5	21.9	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1268	18.6	1193	5 Q9Y1X8	Q9Y1X8 ephydactia f
18	1180.5	17.3	1368	5 Q23821	Q23821 caenorhabdi
19	1142	16.8	1217	5 Q26566	Q26566 schistosoma
20	1126	16.5	527	13 Q90836	Q90836 gallus galli
21	1001.5	14.7	478	11 Q9ESE0	Q9ESE0 ratcus nov
22	942.5	13.8	599	13 Q9PSH2	Q9PSH2 gallus galli
23	906	13.3	165	4 Q14256	Q14256 homo sapien
24	887	13.0	176	11 Q923V5	Q923V5 ratcus nov
25	806.5	11.8	346	13 P11776	P11776 xiphophorus
26	778	11.4	435	5 Q8SZW1	Q8SZW1 drosophila
27	754.5	11.1	311	13 Q99162	Q99162 xiphophorus
28	741.5	10.9	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
29	734	10.8	331	4 Q9BDU7	Q9BDU7 homo sapien
30	725	10.6	1671	5 Q9NTJ5	Q9NTJ5 blomphalari
31	723	10.6	149	6 Q9BG66	Q9BG66 oryctolagus
32	680.5	10.0	1368	13 Q8UW85	Q8UW85 paratichthy
33	680	10.0	1418	13 Q93457	Q93457 scophthalmu
34	673.5	9.9	1369	13 Q8UW86	Q8UW86 paratichthy
35	672	9.9	1412	13 Q8UW84	Q8UW84 paratichthy
36	660.5	9.7	1472	5 Q9U5A8	Q9U5A8 bombyx mori
37	658	9.7	1358	13 Q73798	Q73798 xenopus lae
38	640	9.4	1245	13 Q9YGR8	Q9YGR8 scophthalmu
39	636.5	9.3	1418	13 Q8UW83	Q8UW83 paratichthy
40	630.5	9.3	2144	5 Q9YD94	Q9YD94 drosophila
41	624	9.2	1371	11 Q9QVW4	Q9QVW4 ratcus sp.
42	602.5	8.8	1091	4 Q9UMQ4	Q9UMQ4 homo sapien
43	595.5	8.7	1072	4 Q9BTR0	Q9BTR0 homo sapien
44	595	8.7	987	11 Q91YW0	Q91YW0 mus musculu
45	586	8.6	1036	4 Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

ID O18735 PRELIMINARY: PRT: 1259 AA.

AC O18735;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DI 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE EMBL-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA2127.1; -
DR HSP; P11362; 1FGK.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR004019; YIP_motif.
DR Pfam; PF00157; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YIP_2.
DR ProDom; PDOM00001; Euk_pkinase; 1.
DR SMART; SMO0261; FU; 3.
DR SMART; SMO0219; Tyrc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4CD46 CRC64;

Query Match 89.5%; Score 6097; DB 6; Length 1259;
 Best Local Similarity 89.1%; Pred. No. 0;
 Matches 1126; Conservative 48; Mismatches 74; Indels 16; Gaps 4;

QY 1 MELALCRWGLLLLLPPGAASVQVCTGTMTKRLPSPSPETHLDMHLYOGGVVQGNL 60
 DB 1 MELALCRWGLLLLLPPGAASVQVCTGTMTKRLPSPSPETHLDMHLYOGGVVQGNL 60
 QY 61 ELTYLPNNAISLFDIOIEVGVYLIANOVROVLIQRLIRVGTQLFEDNYALAVLDNG 120
 DB 61 ELTYLPNNAISLFDIOIEVGVYLIANOVROVLIQRLIRVGTQLFEDNYALAVLDNG 120
 QY 121 DLNNTTPTVTSAGSGRLRLSLTELKGGVLIQGNPOLCYODTILMDIFHKNQOLA 180
 DB 121 DLNNTTPTVTSAGSGRLRLSLTELKGGVLIQGNPOLCYODTILMDIFHKNQOLA 180
 QY 121 DPLEGGIAPGAAGGRLRLSLTELKGGVLIQGNPOLCYODTILMDIFHKNQOLA 180
 DB 121 DPLEGGIAPGAAGGRLRLSLTELKGGVLIQGNPOLCYODTILMDIFHKNQOLA 180
 QY 181 LTLIDTNRSPACHPCSPWCKGRCWGESSEDDCSLRTVCGAGCARCKGRLPTDCHEQC 240
 DB 181 LTLIDTNRSPACHPCSPWCKGRCWGESSEDDCSLRTVCGAGCARCKGRLPTDCHEQC 240
 QY 241 AAGCTGPHNSDCLACIHNHSGICELHCPALVYNTDTFESMPREGRYTFGASCYTAAC 300
 DB 241 AAGCTGPHNSDCLACIHNHSGICELHCPALVYNTDTFESMPREGRYTFGASCYTAAC 300
 QY 301 YNYLSTDVGSCTIACPLHNOEVTAEADGTORCEKSKPCARCYGLGMQYIKANKFTGIT 360
 DB 301 YNYLSTDVGSCTIACPLHNOEVTAEADGTORCEKSKPCARCYGLGMQYIKANKFTGIT 360
 QY 361 ELERAGCKKITGSLAFLEPESFDGPASTAPLOEQLOVFTLEETIGYLYISAMPDLP 420
 DB 361 ELERAGCKKITGSLAFLEPESFDGPASTAPLOEQLOVFTLEETIGYLYISAMPDLP 420
 QY 421 DLSEFONLOVIRGRIHNGAYSILLOGLIGISMLGRLSRELGSGLALIHNTHLCPFTV 480
 DB 421 DLSEFONLOVIRGRIHNGAYSILLOGLIGISMLGRLSRELGSGLALIHNTHLCPFTV 480
 QY 481 PWDOLFRRNPHQALLHSANRPEECVGEGLACYP-CAGHCHCGMPPTCCVNSQPLRQEC 539
 DB 481 PWDOLFRRNPHQALLHSANRPEECVGEGLACYP-CAGHCHCGMPPTCCVNSQPLRQEC 539
 QY 541 VEEGRVLOGLPREVYNAHCLPCHPECQPNQSVTCGPEPDQVACAHYNDPFCYARC 600
 DB 541 VEEGRVLOGLPREVYNAHCLPCHPECQPNQSVTCGPEPDQVACAHYNDPFCYARC 600
 QY 601 PSYKPPDLSTYMPIMKPEDEBACQPCPIINCHSCVDDDKCAPEQASPLSTISAVG 660
 DB 601 PSYKPPDLSTYMPIMKPEDEBACQPCPIINCHSCVDDDKCAPEQASPLSTISAVG 660
 QY 660 ILLAVVGLVGLIKRRRQKIKRTMRLLQSTEL--VEPLPSGAMFNQOMRILK 715
 DB 660 ILLAVVGLVGLIKRRRQKIKRTMRLLQSTEL--VEPLPSGAMFNQOMRILK 715
 QY 715 ETELRKVKVLSGAGFYKGIWIPDGENVKIPVAIKYLRENTSPKANKETLDAVYVAG 774
 DB 715 ETELRKVKVLSGAGFYKGIWIPDGENVKIPVAIKYLRENTSPKANKETLDAVYVAG 774
 QY 775 VGSPPVSKLLGICLTSTVQVLTQMLPVCCLLDHRENGRIGSODILNMCQIAKMGSYL 834
 DB 775 VGSPPVSKLLGICLTSTVQVLTQMLPVCCLLDHRENGRIGSODILNMCQIAKMGSYL 834
 QY 835 EDVLRVHRDLAARNLVKSPNHVKITDGLARLLDIBTEHAAGKVPIMVMALESILR 894
 DB 835 EDVLRVHRDLAARNLVKSPNHVKITDGLARLLDIBTEHAAGKVPIMVMALESILR 894
 QY 895 RRFTHQSDVMSYGVTVWELMTFGAKPYGIGAREIPDLLEGESELPPOPTITDIVYIMV 954
 DB 895 RRFTHQSDVMSYGVTVWELMTFGAKPYGIGAREIPDLLEGESELPPOPTITDIVYIMV 954
 QY 955 KCMIDSECRPREFRELVESEFARMARDPQRFVYIQNEDGSPASPLDSTFYRSLLDDDDMGD 1014
 DB 955 KCMIDSECRPREFRELVESEFARMARDPQRFVYIQNEDGSPASPLDSTFYRSLLDDDDMGD 1014
 QY 1015 LVDABEYLVPOQGFPCPDPAFAGAGMTHRRSSSTRSGGSDTLTLGLEPSEEEARPSPLA 1074
 DB 1015 LVDABEYLVPOQGFPCPDPAFAGAGMTHRRSSSTRSGGSDTLTLGLEPSEEEARPSPLA 1074

DB 1016 LVDABEYLVPOQGFPCPDPAFAGAGMTHRRSSSTRSGGSDTLTLGLEPSEEEARPSPLA 1075
 QY 1075 PSEGASDVFDDDLGMGAAGKQSLPTHDPSFLQRYSDPTVPLPSETDGVYAVLTCSPQ 1134
 DB 1076 PSEGASDVFDDDLGMGAAGKQSLPTHDPSFLQRYSDPTVPLPSETDGVYAVLTCSPQ 1135
 QY 1135 PEYVNOPEVPPQPPALBEPPLPSPRAGATLERPKTLSPKTLSPGNVGVKDFAFGSAV 1195
 DB 1136 PEYVNOPEVPPQPPALBEPPLPSPRAGATLERPKTLSPKTLSPGNVGVKDFAFGSAV 1195
 QY 1190 ENPEYLTPOGGAAPOPHPPEAFSPAFDNLYYWDOPPERGAPSTFTGPTAENPEYLT 1249
 DB 1196 ENPEYLTPOGGAAPOPHPPEAFSPAFDNLYYWDOPPERGAPSTFTGPTAENPEYLT 1255
 QY 1250 DVPV 1253
 DB 1256 DVPV 1259

RESULT 2

Q9QX70 ID Q9QX70 PRELIMINARY; PRT; 1209 AA.

AC Q9QX70; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Epidermal growth factor receptor.

GN EGFR.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_Taxid=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=FISHER; TISSUE=LIVER;

RX MEDLINE=90258888; PubMed=2342466;

RA Patch L.A.; Harris J.; Raymond V.W.; Blaband A.J.; Lee D.C.;

RA Patch H.S.;

RT "A truncated, secreted form of the epidermal growth factor receptor is

RT encoded by an alternatively spliced transcript in normal rat tissue.";

RL Mol. Cell. Biol. 10:2973-2982 (1990).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=FISHER; TISSUE=LIVER;

RA Guttridge K.; Dawson T.L.; Eard H.S.;

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

EMBL; M37394; AAF14008.1; -

HSSP; P11362; 1FGK.

DR InterPro: IPR000494; EGFR L domain.

DR InterPro: IPR000719; Euk_Pkinase.

DR InterPro: IPR002174; Furin-like.

DR InterPro: IPR001245; Tyr_Pkinase.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF00069; Pkinase; 1.

DR Pfam: PF01030; Recept L domain; 2.

DR PRINTS: PR00109; TYRKINASE.

DR ProDom: PDD00001; Euk_Pkinase; 1.

DR SMART: SM00219; TyrKc; 1.

DR SMART: SM00107; PROTEIN KINASE_ATP; 1.

DR PROSITE: PS00111; PROTEIN KINASE_DOM; 1.

DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.

KM ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.

SO SEQUENCE 1209 AA; 134891 MW; 96FE87FECB1B7773 CRC64;

Query Match 44.9%; Score 3060; DB 11; Length 1209;
 Best Local Similarity 49.3%; Pred. No. 8; 9e-223;

Matches 630; Conservative 164; Mismatches 366; Indels 118; Gaps 28;

QY 3 LAALCRWGLIALLPAGA-ASTOYCTGTDMKRLPAPERTHMLRLHYGGVGNNE 61
 DB 15 LAALCAAG-----GALBEKVCQGTNRKLTQLGFEDHFSLQMGFNCCVVLGNLE 66
 QY 62 LTYLPTNASLSFLQDIOEVQGVYLIANQVRQVPLQRLIRVSTQLFEDNYALAVLNDG 121
 DB 67 ITVQRNYDLSFLKTIQEVAGVYLIANTVERIPLLELIQIRGNALYENTYALAVLSN-- 124
 QY 122 PLNNTPTVGTASPEGJRELQRLSTELTKGVLIQRPOLCYODTLMKDIFKXNQAL 181
 DB 125 -----YGTNKTGLBELPRLNQLLIGAVRSPNNPILCNMTIQMRDLY-QDVPLS 175
 QY 182 TLIDYTRS-RACHPCSPCKSGSRGWSSESDCCSLTRTYCAGCA-RCKGPLEPTDCHEQ 239
 DB 176 MSMDVQRHLTGCRKCDPSCPNGSCWGRGGENCQKLTITICAGCCSRRCGRSPSDCHNQ 235
 QY 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALTYNTDTESMPNPEGRTFRASCYTAC 299
 DB 236 CAAGCTGPRSDCLVCHRFDEATCKTCPLMLYNTTYQMDVNEBGKXSFSGATCVKXC 295
 QY 300 PNYLSTDVGSCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGI 359
 DB 296 PRNVVTDHSGCYRACGPDYEV-EDGVGSKCKKCGKPCXKNGIGISEFK-DTLSINA 353
 QY 360 TELE-FAGCKKITGSLFLPESPDGPASTAPLOEQLOVFTLEITGYLYISMPDS 418
 DB 354 TIKHFKYCTAISGDILPLFAFKGSPFRTPPLDPELEILITVKEITGFLLIQAMPEN 413
 QY 419 LPDLSTVFQNTQVTRGRILHNGAVSLTQGLISMLGRSLREKSGALIHHTHCIFYH 478
 DB 414 WTDLHAFENLEITRGKQHGQSLAVAGLNTSLGRSKETISDGVIIISGRNLCYAH 473
 QY 479 TVPMDQLFRNPHQALLHTANRPDECEVBEGLACHOLCARHCWPGPTQCVNCSQPLRQ 538
 DB 474 TIKWKLFGTPNCKTKIMNRAEKDCATNHCNPLCSSBSCGPEPTDCVSGQVNSRGR 533
 QY 539 ECEVEGRVLOGLPREYVNAHCLPCHPECPONGSVTCPEFADQCAVACHYDPRPCVA 598
 DB 534 ECVDCNLTGEBREFEENSECQCHCECPOTMNTTCGRGDNCKCKAHYVDGHCYK 593
 QY 599 RCPGSKVPELSTYPIWKFPDEBAGCQPCPINTHSCVDLDDKCPAQRASP-LTSIVA 657
 DB 594 TCSSGIMGENNTL-VWKFADANNVCHLCHANCTYGCAGPGLKCC-QQPEGRPIELIANG 650
 QY 658 VNGILLVVUVGVVFGILIF--NNFTVSPFLRPKYSASHLEPLTSGAMPNOQMRILK 714
 DB 651 IVGGLLFIIV-VALGIGLFRRRQDLVAKRTLRLLQERLEVEPLTPSGEAPNQAHRLIK 709
 QY 715 ETELKRVKLVGSAFGTVYKGIWIPGENYKIFVAIKVLRENTSPANKETIDEAYVMAG 774
 DB 710 ETEFKTKLVGSAFGTVYKGLMIPGEKXKIPVALIKELREALSPANKETIDEAYVMS 769
 QY 775 VSGPYVSRLLIGTCTSTVOLVTLMPYGLLDHVRENRGLSGODLLNMCQIAKMSYL 834
 DB 770 VDMPHVCRLLIGTCTSTVOLITLMPYGLLDYVRSHKNDIGSYLLNMCVQIAKMNLY 829
 QY 835 EDVRLVARDLAANVLKSPNHVKITDPGLARLLDIDETPYHNDGKVPILKXMALESLIR 894
 DB 830 EDRRLVHRDLAANVLKTPQHKITDFGLAKLIGBEKRYHAEGKVPILKXMALESLIR 889
 QY 895 RRTTHOSDWSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDYMINV 954
 DB 890 RIYTHOSDWSYGVYTWELMTFGSKPYDGIIPASEISILEKGERLPQPPICITDYMINV 949
 QY 955 KCMWIDSECPREPRVSESRMARDPQRFVVO-NEDGSPASPLDSTYRSLLBEDDDMG 1013
 DB 950 KCMWIDSECPREPRVSESRMARDPQRFVVO-NEDGSPASPLDSTYRSLLBEDDDMG 1009
 QY 1014 DIVDAEEYLVPOGGFFCPDPAPGAGMVHHRSSSTRSGGDLTLGLPSEBEAPRSL 1073
 DB 1010 DIVDADEYLVPOGGFF-----NSPST-----SRTPL 1035

QY 1074 APEAGSPVFDGDLGMAKAGLQSLPTHPDPELQRYSEDPVTPVPSER--DGYVAPLTC 1131
 DB 1036 LSSLANSN-----SSTVACINRNGSCRYKEDFLQRYSDPFSVLTEDNIDDTFL----- 1086
 QY 1132 SPOPEYVQPDVPPPPSPREGPLPAPAPAGATLERAKTLSPGKNGVAVKVAFGAVEN 1191
 DB 1087 -PVPEYING-SVPKRPAGSVQNPVYHNPPLPH-----ABGRDLHYQN--PHSNVAVSN 1134
 QY 1192 PEYL-TPQGAAPQCHPPAPSPAFDNLVWPO-----DP-----PERGAPST 1234
 DB 1135 PEYLNTAQ-----PTCLSSGFDSSALMTQKSHQMSLDNPYQDFFPKKXPNGI 1185
 QY 1235 FKGTPTAENPEYGLDVP 1252
 DB 1186 FKQ-PTAENPEYLRVAPP 1202

RESULT 3
 Q9EP98
 ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
 AC Q9EP98;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor isoform 1.
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balseubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maithe N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schell C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balseubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maithe N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275366; AAG28045.1; -
 DR EMBL; AF275364; AAG28045.1; JOINED.
 DR EMBL; AF275365; AAG28045.1; JOINED.
 DR EMBL; AF275367; AAG24386.1; -
 DR HSSP; p11362; 1FGK.
 DR MGD; MGI:95294; Egfr.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; Fu; 5.
 DR SMART; SM00220; S_TKC; 1.
 DR SMART; SM00219; TyrKc; 1.

DR PROSITE, PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE, PS00107; PROTEIN KINASE, ATP, 1.
 DR PROSITE, PS00101; PROTEIN KINASE, DOM, 1.
 DR PROSITE, PS00109; PROTEIN KINASE, TYR, 1.
 DR ATP-binding; Receptor; Transferase.
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD0021C9DE32E18 CRC64;

Query Match 44.6%; Score 3038; DB 11; Length 1210;
 Best Local Similarity 48.9%; Pred. No. 4,1e-221;
 Matches 622; Conservative 167; Mismatches 370; Indels 114; Gaps 25;

11 LLLALLPFGAA--STOVCTGDMKRLPASPEHLDMLRLHLYOCQVQVQGLLETTYPTN 68
 14 LTLALCAAGALEKCKVCGTSNRLTQGTEDHFLQRYNNCEVVLGNLETTYQRY 73
 69 ASLSFLDIOEVQGVILAHNQVROVLIQRLIYRGQLFEDNYALVLDNGDPLNTP 128
 74 YLSFLKTIQEVAGVILALNTVERIPLENLIIRGNALYENTYALALISN----- 124
 129 VTGASPGGLRELQRLSLTEILKGGVLIQRIQNPQLCYOPTILMKDI----FHKNNQALTLI 184
 125 -YGINRTGLRELKRNLDLIGAVRSNNPILCNMTLQWRDIYQVNFPSNMSMDL--- 180
 185 DTNRSRACHPCSPMCKGSRGWSSSDCQSLTRTVCAAGCA-RCKGRLPTDCHEQCAAG 243
 181 -QSHPSGCPKCDPCSPGSCWGGGHEHCQKLTIKIICAQCSHRCGRSSPSDCCHQCAG 239
 244 CTGPKHSDCLALPHNHSGLICELHCPALVTYNTDPESMPEGRYTFGASCVACAPNY 303
 240 CTGPRESCLVQCKFQDEATCKDTPPLMLYNTPTQDVNBERKYSAGATCKVCKCRNY 299
 304 LSTVSGCTLVCPHNDVTAEDGTORCEKSKPCARCYGLAQVYIKANSKFIGITELE 363
 300 VVTHGSCVRAACGPDYEV-EDGIRKCKKCDGCRKCNIGIGEFK-DTLSINATNIK 357
 364 -FAGCKTFGSLAPLPSFDDPDASNTAPLOPELOVPELLETTGYLYISAWDSLPL 422
 358 HFKYCTAISGDLHLIPVAFKGSDFRPLDPRELEIKTYKELTGILLQAWDNTDL 417
 423 SVFQMLQYIRGLIHNGAVSLTLOGLISMLGSLRLSGLSLIHNTLHCVHTVPW 482
 418 HAFENLEIRGRTHQGFSLAVGLNTSLGRLSKISGDIISGNRLCYANITNW 477
 483 DQFRNPQALHTANPEDECVGEGLAGHOLCARGHWGPGPQCVCNCSOFLRGQCEVE 542
 478 KKLFGTPQKTKIKNNRAEKXCKKAVNHNCPNLCSEGGGEPDVCSCQVNSGRECEVE 537
 543 ECRVQGLPREYVNAARHCLPCHECOPONGSVTCFGEPADQCVACAHYKDPFCVACPS 602
 538 KCNILEGEPREFEVSESCIQCHPECLPQANNITCTGRGPDNCLCAHAHIDGPHCVKCPA 597
 603 GVKPDLSTYPIWKPFDEBGACQPCPINCTHSCVDLDKGCFAEGRASPLTISVAVVGL 662
 598 GIMGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLGCVWSPGKIPISATGIQVGL 656
 663 LVVYGVVFGILIF---NNFTVSPFLRVKVSASHLEPLTPSGAMPNQAORILKETELR 719
 657 LFIIV-VALGIGLFRKRRHIVKRLRLLEBELVEPLTSGEAPNDAHRIKETEPRK 715
 720 KVKVLGSGAFGVYKGINPDGENYKIPVAIKVIRENTSPRANKELIDEAVYMGASVY 779
 716 KIKVLGSGAFGVYKGLWIPGEKVKIPVAIKELREATSPRANKELIDEAVYMGASVY 775
 780 VSRLLGICLTSTVOLVTLQMPYGLLDHVRNBSGLSGODLNNCMOIAKMSYLEVRL 839
 776 VCRLLGICLTSTVOLVTLQMPYGLLDHVRNBSGLSGODLNNCMOIAKMSYLEVRL 835
 840 VHRDLAANVIVKSPDNHVKITDFGLARLLDIDETEHADGSKVPIKMALESILRRFTH 899
 836 VHRDLAANVIVKTPQHKITDFGLAKLGAEEKEVHLEGKVPKMALESILHRIYTH 895
 900 OSDVWSYGVTVWELMTFGAKPYDGI-PAREIDPLEKGRRLQOPICITIDVYMWVKMMI 959

DB 896 OSDVWSYGVTVWELMTFGSKPYDGI-PASDISLLEKGRLEPOPICTIDVYMWVKMMI 955
 QY 960 DECEPRFRELVSFSRMARDQRFVVIQ-NEDLCAPASPLSTFTRSLLEDMDGLDYDA 1018
 DB 956 DADSPRFRELILLESKMARDDQRLVIVQGDGRHMLPSPTDNPFRALMDEEDMDYVDA 1015
 QY 1019 EBYLVQGGFPCFCDPAPAGAGVHHRRHSSSTRSGGDLTLLEPSEEPAPSPAPSEG 1078
 DB 1016 DEYLVFQGGF-----NSPST-----STPILLSLS 1041
 QY 1079 AGSDVFDGLGMAAGKGLSLPTDPSPLQRYSEDPVLPSET-DGVAPLTSQPOE 1136
 DB 1042 ATSN-----NSTVACINRNSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----PVBE 1091
 QY 1137 YVNPDPYRQPPSPREGPLPAPRPAGATLIERAKTISPGKNGYVXVAFAGAVENPEYL 1195
 DB 1092 YVNO-SVPRPAGSVQNPVYHNPQPLHP-----APGDLHYQN-PRSNVAGNEEYLN 1140
 QY 1196 TPQGAAPQPPPPAPSPAFDNLVYWDQ-----DP-----PERGAPSPSTFKGTP 1239
 DB 1141 TAO-----PTCLSSGFSNPALWQKSHQMSLDNDYQGFPPKTKENGIFKG-P 1190
 QY 1240 TAENPEYLGIDVP 1252
 DB 1191 TAENAEYLRVAP 1203

RESULT 4
 Q9YH40 PRELIMINARY; PRT; 1165 AA.
 ID Q9YH40
 AC Q9YH40;
 DT 01-MAY-1999 (TREMELREL. 10, Created)
 DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XHRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 OC Acanthomorphia; Acanthopterygii; Percomorphia; Atherinomorphia;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 NX NCBI_TaxID=8086;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
 RA Altschmidt J., Scharl M.;
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 overexpression and mutational alterations.";
 RU Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA STRAIN=RIO PURIFICATION;
 RA Scharl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U53471; AAD10500.2; -;
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000345; Cyt_c_heme_bind.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00669; pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; Fy; 3.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN_1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.

DR PROSITE: P500107; PROTEIN KINASE ATP; 1.
 DR PROSITE: P550011; PROTEIN KINASE DOM; 2.
 DR PROSITE: P500109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7f7EB38D6771A74E CRC64;

Query Match 38.8%; Score 2647; DB 13; Length 1165;
 Best Local Similarity 44.6%; Pred. No. 1,7e-191;
 Matches 571; Conservative 164; Mismatches 388; Indels 156; Gaps 32;

QY 1 MELALCRMGILLALLPRG-AAST---CYCTGDMKLRPASBETHLMDRLHYGCGY 55
 DB 4 LLEL---LLELLISGRCCSTDPDRKVCQGTSGNMTM---LDNHYLKKKKKYGCCNV 56
 QY 56 VQGNLELYLPTNLSFLDIOEVQGVLIANQVQVPLQRLIRYGTQLFEDYVALA 115
 DB 57 VLENLEITTOENDDLSFLOSIGVGVLIAANEVSTIPLVNLRLRGONLYEGNFTLL 116
 QY 116 VLDNGDPLNNTPTVTGASFGGLRELQSLSTELIKGVLIQNPQLCYQDTIMKDFHK 175
 DB 117 VMSVYOK-NPSSP--DYYQVGLKQLSLNTEILSGGVKVNHPLLCNVETIMWDIVDK 173
 QY 176 NNQALTLIDTNRBACHPCSPMKGSRGWSESDCQSLTETVCAGGC-ARCKGPLPTD 234
 DB 174 TSNPTMNLIPHAERQCKDPCGVNASCNAPRGHCQKTKLCAEQCNRRKGRKPID 233
 QY 235 CCHQCAAGCTGPRHSDCLALHNHSGICELHCPALVYNTDTFESMPNDEGRYTFGAS 294
 DB 234 CCHHCAAGCTGPRATDCLACRPFNDGCTKTPPKKIDIVSHQVNDPNKXTEFGAA 293
 QY 295 CVAACPVNLTSTVGSCTVCPLNCEVTAEDEGQREKSKCAVYCLGM----QYI 350
 DB 294 CVECEGPNVYVTE-GACVRSCKGMLVD-ENGKRSKPDGVCPRKCDIGISLNTI 351
 QY 351 KANSKFIGITELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOEQLVFEITEITGYL 410
 DB 352 AVNSTNG----SESNCTKINGDIILNRNSFEDGPHKIGPMPENHMLTYKELTGYL 407
 QY 411 YISAMPDSDLDSVFQMLQYIRGILHNGAYS-LTQGLGISMLGRSLRELGSGLALH 469
 DB 408 VIMWPEMNTSLSVFQMLIIRGRTTFSGRFSFVVQVSHLQWGLSLKEVAGVILK 467
 QY 470 HNTHLCFHTVPMQDLFRNPHQALHMTANRPEDECEGAGACQQLCARGCMGPGTQCV 529
 DB 468 NTPQLRKASTINRRRLPRSEDQIEVDART-----ENQTCNNEGSEDCMGPGTMCV 520
 QY 530 NGSQFLRGCEVEECRYLQGLPREYVNAHCLPCHPBCQPFQNGSVTCFGEADQCYAAC 589
 DB 521 SCHLVDRGRCVASCMLQGEPRERAOVDGRVCVQCHQECVQTSILTCYGGPANCSKCAH 580
 QY 590 YKDPFCVARGCPGSKVDLSYMPIMKPEDEGACQPCPICTHSCVVDLDDKGRPAQRAS 649
 DB 581 FQDQPCQICPRCPHMLGSDPTL-TWKADMGQCCQCHQNTGCCGPGISGGRCD-IYS 638
 QY 650 PLTISVAVVGIILVVLGVVFIL-----IFNNFVSEFWLRVPKVASHLBELTPSGAM 704
 DB 639 HSSLAVGLVSGLLITVALLIVLLRRRIKRRKTIIRLLQCKE---VEPLTSGQA 694
 QY 705 PNAQWRILKETELRKVKYLSGAFGVVYKGIWIPGGENVKIPVALIKYLRNTSPKANE 764
 DB 695 PNAFLRLKETELFKDQRYLSGAFGVVYKGIWIPGGENVKIPVALIKYLRNTSPKANE 754
 QY 765 ILDEAVYMGVGVSRLLIGLCTSTVQVLTQVLMPEYGLLDHRENRGLSGQDLNAC 824
 DB 755 VLDEAVYMGVGVSRLLIGLCTSTVQVLTQVLMPEYGLLDHRENRGLSGQDLNAC 814
 QY 825 MQIAKMSYLEDVRLVHRLDAANVLYKSPNHKIDPGLARLLDIDETIHAAGSKVPI 884
 DB 815 VOIAKMSYLEDVRLVHRLDAANVLYKSPNHKIDPGLARLLDIDETIHAAGSKVPI 874
 QY 885 KMALESILRRFTHOSDVMSYGVYWEMLTFAKPYDGIAPAEIDLEKEGRDLQPPPI 944
 DB 875 KMALESILQWYTHOSDVMSYGVYWEMLTFAKPYDGIAPAEIDLEKEGRDLQPPPI 934

QY 945 CTIDVYMLWKCMIDSECRPRFRELVSFSMARDPQRFVYIYONEDLGPASPLDSTFVR 1004
 DB 935 CTIEYMTMLKCMIDSPSRPRFRELVSFSMARDPKSYLVYIQ--NLPSPSDRLLS 991
 QY 1005 SLEDDMDKDLVDAEYVLPQGFCCPPAPAGAGVWHRRSSSTRSGGDLTGLSPS 1064
 DB 992 RLSSSD--DVADADBYLL-----RYKRN-RQSS----- 1018
 QY 1065 EEEARPSPLAEEAGAGSVDFGDLGKAQKLSLPTHPDPLQRYSEDPV-PLPSETD 1123
 DB 1019 -----EPGIPNGH-----FVRENSIALRYISDPTOMALEKLD 1052
 QY 1124 GYVAPLTCSPQPEYVNPDPVPPQ-----PSPRE-----GPLP-AAPACATLERAKT 1170
 DB 1053 GH-----EYVNPQSESTSSRLSDIYNVYEDLIDGMQPVLSLSEALTNFSRPY 1102
 QY 1171 LSPGKGVKVDYFAFGAVENDEYLTPOGGAAPQHPPEAFSPAFDNLVYWDQDPPERGA 1230
 DB 1103 LNTNONSLL--PLVSSGSVDPDY--QAG-----YQAAF-----LPQTGA 1137
 QY 1231 PSTFKGPTPAENPEYLG 1249
 DB 1138 LTGNGMFLPAENLEYLGL 1156

RESULT 5
 Q9W6F6 PRELIMINARY; PRT: 1137 AA.
 AC Q9W6F6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN ERB4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC TISSUE=HINDRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.;
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
 RT embryonic chick hindbrain";
 RL Mol. Cell. Neurosci. 13:237-258 (1999).
 DR EMBL: AF121963; AAD31764.1; --
 DR HSSP: P11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001368; TNFR_C6.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF01030; Recep_L domain; 1.
 DR Pfam: PF02157; YLP; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SMO0261; Fu; 3.
 DR SMART: SMO0219; Tyrc; 1.
 DR PROSITE: P500107; PROTEIN KINASE ATP; 1.
 DR PROSITE: P550011; PROTEIN KINASE DOM; 1.
 DR PROSITE: P500109; PROTEIN KINASE TYR; 1.
 DR PROSITE: P500652; TNFR_NGFR_1; UNKNOWN_1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match

38.7%; Score 2636.5; DB 13; Length 1137;

Best Local Similarity 45.9%; Pred. No. 1e-190;
Matches 525; Conservative 172; Mismatches 357; Indels 91; Gaps 26;

161 LCYODTLMKDIFFKNNQALTLIDTRSPACHPCSPCKSGRCWSSSEDCSLTFVTC 220
3 LCFADTLMHODIVANPNASNTLIVPTNGSSCCGRCHKSCGT-RMGFTENHCOTLTCTVC 61

221 AGGC-ARCKGFLPDDCCHEQCACTGPKSDDLACIHFHNSGICELHGCALVTYNDTF 279
62 AECCDGRCYGYVADCCHECAGCGSKPDTDCFACNCFNDSCGCTQCQTFVYNTTF 121

280 ESMNPBGRYTFGASCYACPYNYLSTDVGSCTVCPJLHNOEVAEDGTORCEKSKPCA 339
122 QLEHNNAKTYGAFCYKCKPHNV-VDSSCVACASSKMEV-EENIGIKCKRPTDICI 179

340 RVCYGLGMOYIKANSKFIQTELE-FAGCKKIFSGSLAFBESFGDPASTAPLPQEQ 398
180 KACDGIQGSF-VSAQTVDSNIDKFNCTKINGNLFLVTGIGHDEPHYTAIINPEKN 238

399 VFETLEETGYLYISAMPDSLDPDSVFQNLQVIRGLIHNGVSLTQGLISWLGIRSL 458
239 IFQVREITGYLNIQSPENMTDRVPSNLVTIGRALYSGSLILKQGITSLQFQSL 298

459 RELSGALIHNNHLCFVHTVMDQLFRRHQALHTANBEDECVSEGLACHQLCARG 518
299 KQISAGNIYITDNSNLCTYHTVNTSLFSTPSQKTVIHRNKAENCTADGVCNELCSSD 358

519 HCMGPPTQCVNCGOPFRCGECEVRCVLOGLPREYVNAHCLPCHPECOP-ONGSVTFC 577
359 CGWGPBQDCLCKRFIRGTCISCNLYDEFEFPANGSCVCMCDQCEMEENMTTCY 418

578 GPEADOCVACAHYDPEFCVACPSGVKPDLSYMPIMKPPDEGACOPCPINCTHSCVDL 637
419 GPGDHCTKCHFPDGPVCEKCPDGLQGANF--IFKYADEDECHCHPCHNCTQGRGP 476

638 DDGCG-----PAEORASPLTSVSAVV-GILLVVLGVVFGILLFNNTVSFWL 685
477 ASHDCIYYPWTRQSTLPQHAR-TPL-IAAGVIGLFTIVMGSLTFAYVYRK----- 526

686 RVPKVSASH-----LEPLTPSGAMPQAOARILKETELRKVYLSGAFVYKGIWI 738
527 SIKKRAIRRFLETETELVPLTPSGTAPQAOARILKETELRKVYLSGAFVYKGIWI 586

739 PDGENVKIPIVAKVURENTSPKANKEILDEYVYVAGVSPVVSLLGICLTSTVQVLTOL 798
587 PEGETVKIPIVAKILNETTGPKANVEFDEALIMASMDPHLVLLGVCLSPITQVLTOL 646

799 MPYGLLDHVENRGRGLSGODLLNMCQIAKMSYLEDVRLVHBDLAARVLYKSPNHVK 858
647 MPHGLLDYVHEHKNIGSOLLNMCQIAKMSYLEDVRLVHBDLAARVLYKSPNHVK 706

859 ITDGLAALLDIDETEHYADGKVPDKMMALBSILRRRPTHQSDVSYGVTVWELMTPGA 918
707 ITDGLAALLDIDETEHYADGKVPDKMMALBSILRRRPTHQSDVSYGVTVWELMTPGA 766

919 KPYDIPAREIPDLLEKGERLPQPICTIDVYIMVCMIDSECRPFRELYEFSMA 978
767 KPYDIPAREIPDLLEKGERLPQPICTIDVYIMVCMIDSECRPFRELYEFSMA 826

979 RDQRFVYIIONED-LGPASPLDSTFYRSLLDEDDMGDLVDAEYLVPOQGFCCDPAPGA 1037
827 RDQRFVYIIONED-LGPASPLDSTFYRSLLDEDDMGDLVDAEYLVPOQGFCCDPAPGA 885

1038 GGMVHHRSSSTRSGGDDLTLGLEPSEBEAPRS--PLAP--SEGSGSVFPGDGLGMAK 1094
886 RTRIDSNKQVYRQGYAABGV-PMYRAPGCIIPAPVQAQATATIFEDTCNGTLR 944

1095 GLQSLPDPSPLOQYSEDFTVPLPS-----ETDGYVAPLTPSPQEVYVQDPVAPQP 1147
945 KOVATLAKESSTQYVSADPTVFIPIERVIRGLDEBDGWTMPMRDKPTDYNPVENPFI 1004

1148 PSPREGPLPAA-RPAGATLIERAKTILSPKNGVYKVF-----AFGGAENREYLLTPQ 1198

DB 1005 SRKXGDDLOAVNDPEYHN-----APNGPAADEYVNEPYLNTFANTLENAEYL--- 1054

QY 1199 GGAAPQHPPEPAPFSAFDNLVYWDODPBERGA--PSPFKGTPT-----AE 1242

DB 1055 -----KNLPERAKKAFADNPDMHNSLPPRSTLQHPDYLOEYSTFYKONGRIRPIVAE 1109

QY 1243 NPEYL 1247

DB 1110 NPEYL 1114

RESULT 6

ID P79754 PRELIMINARY; PRT: 1328 AA.

AC P79754;

DT 01-MAY-1997 (TEMBLrel. 03, Created)

DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)

DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

DE Erib3.

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;

OC Tetraodontidae; Takifugu.

OX NCBI_TaxID=31033;

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE=9917347; PubMed=1007531;

RA Gellner K., Brenner S.

RT "Analysis of 148 kb of genomic DNA around the whl1 locus of Fugu rubripes."

RL Genome Res. 9:251-258(1999).

DR EMBL; AF056116; AAC34391.1; -

DR HSSP; P11362; IFGK.

DR InterPro; IPR000494; EGFR_L_domain.

DR InterPro; IPR002174; Euk_Pkinase.

DR InterPro; IPR001245; Tyr_Pkinase.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Pkinase; 1.

DR Pfam; PF01030; Recept_L_domain; 2.

DR Prodom; PD000001; Euk_Pkinase; 1.

DR SMART; SM00261; Ftr; 3.

DR SMART; SM00219; Tyrc; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.

KW ATP-binding; Transferase.

SO SEQUENCE 1328 AA; 148613 MW; A33039258864789 CRC64;

Query Match 33.2%; Score 2265; DB 13; Length 1328;
Best Local Similarity 39.8%; Pred. No. 1.8e-162; Indels 198; Gaps 33;
Matches 513; Conservative 160; Mismatches 418;

9 WGLLLALPP--GAASDQ-----VCTGDMKRLRPASPTHLDMLRLHYGQCVVQGNLEL 62
4 WRLIMCVASRLAAASQGEAVCPETQNGLSGQENQVNLKORYGCEIIMNTLEI 63

QY 63 TYPTASLSFLDIDIEVQGYLIANNQRYVPLQRLRYRGQLFEDVYALAYDNGNP 122
64 TQESWWSFLLTIREYGVYLIANNHFOEILPGLQRLRYRGSILERSFALSFLN--- 120

DB 123 LNTTPTVYSGPGLRELOLRSLTEILKGVLIQRPOLCYOPTILMKDIFHNQNALAT 182
121 ----YKDG--PSGLNQGLMNTLEILDGVOGINNKYLRYGMYVYWRIL--RNMDAPLE 173

QY 183 LIDTNSRACHPCSPMCKSGRCWSSSEDCQSLTRVYKAGGC-ARCKGFLPDDCCHEQCA 241
174 IQNGERGVCH--KSC-GNYCWGPKDCCQILTKTVCAQCQCDRCFGSPRCCHIDEA 229

QY 242 AGCTGPHSDCLACLFHNSGICELCPALVTYNTDFTFSMPYRPEGRTFGASCYACPY 301
230 AGCKGFLDTCFACRCLFNSGACVPCQCPQTLINNKQTFQMETNPNAKYQGSICVSGCCT 289

```

QY 302 NYLSTVYSGCTVLPNLNOETADGTQR--CEKSKRCAVYCYGLAQWYIKANSKFIGIT 360
Db 290 HFV-VDSGSCVSVCPPKMEV--ERGSQRCELCSSGICPVCESTGE----QRQYDSS 342
QY 361 ELE-FACCKKIFGSLAFPEFSDGDPASNTAPLOPELOQVETLEITGYLYISAMPDSL 419
Db 343 NIDSFINCTKIQGSLHFLVGLGDDDEKPNPFLDAKLEVFRTREITDILNTQSWPKEL 402
QY 420 PDLVSFQNLQVIRRIILHNGAYSLTLQGLISWGLSRLSELGSLALHNNHLCVHT 479
Db 403 NDLSVFSSLLTTIQGRSLFKRPSLMMWRIPLTLSIGLSLSEISDGSVYISQNAHLCTHHT 462
QY 480 VPMDOLEFRNPH-QALLHTANRPEDECVEGELACHOLCARHCWMPGPPTQCVNSQPLRG 538
Db 463 VMTQLEFRGSVRANSINSRPMACVADGRVCDPLCSGSCWMPGPDQCLSRNYSRHG 522
QY 539 ECVEECVQLQGLPREYNAH--CLPCHPEQOPQNGSTTCGPREDDQVCAAHKXDPFCV 557
Db 523 TVAGCHFNNGIPREFAGLNGVCAHCECKPQTKASCTGPGADBECACTKTRDGYCM 582
QY 598 ARCPGVPDLISVPIKPKFDEGACOPCPINCTHSCVDLDDKCPAEPASPLTISVA 657
Db 583 SSCPAGVN-DSKXLLFKPRYREHCEPCHQNCTQSGSGPLNDC--LEAHLTLTSSG 638
QY 658 VVGLLVVTVGVVGLIFENFTYSFWLRPKV-----SASHLEPLTPSGAMPN 706
Db 639 ITGIALVPAGLICLVLF--FLQMLYHRLGLAIRKRAMRYLESSEGFELP-GEKGT 695
QY 707 QAKRILKETELRKYKVLGSGAFSTYKGIWIPGDEWVKIPVAIKVIRENTSPKANEIL 766
Db 696 KYHARILKPSDLKIKPLGSGVFSTVSKGFWIPGEVYKIPVAIKTIQDSSGQTFEIT 755
QY 767 DEAYVAVGSPVYSRLIGLITSTVOLVQLMRYGCLLDHVENRGLSGODLLNMCQ 826
Db 756 DHLMSLSLHPYIVRLIGICPCTCLVQLTQLSHGSLLEHIRHKTSLDPPQRLNMCVQ 815
QY 827 IAKMSYLEDVRLVRLHRLAARNLVKSPNHVKITPDLALDLDEHYHADGKXVITK 886
Db 816 IAKMSYLEDVRLVRLHRLAARNLVKSPNHVKITPDLALDLDEHYHADGKXVITK 886
QY 887 MALESILRRRTTHQSDVSYGVTVLMTFGAKPYDIPAREIPDLLEKGERLPORPICT 946
Db 876 MALESILRRRTTHQSDVSYGVTVLMTFGAKPYDIPAREIPDLLEKGERLPORPICT 946
QY 947 IDVYIMVKKCMIDSECRPRRELVSSEFARMADPQRFVITQNEIDLSPASPLDSTFRL 1006
Db 936 IDVYIMVKKCMIDSECRPRRELVSSEFARMADPQRFVITQNEIDLSPASPLDSTFRL 1006
QY 1007 LEDDDMGDLVDAAEYVLPQOGFCPCDPAFGAGVYHHRHSSSTRSGGDLTLGLEPSEE 1066
Db 981 LEDDDMGDLVDAAEYVLPQOGFCPCDPAFGAGVYHHRHSSSTRSGGDLTLGLEPSEE 1066
QY 1067 EAPRSPPLAPSEAGSDVFDGDLGNG--AAKGLQSLTHDPSPIQ-----RYSED 1113
Db 1008 E-----GGRFATPSPLOPSPSWSSTSPQINSYMMWTQLRD-- 1044
QY 1114 PTVLEPSETDGVVAPLTCSPQ--EYVNO-----PVRPQPSPREGPI--P 1156
Db 1045 ----FAVSGGHITGLYLMSPSPVDTIRQLMWQRSLSSVATLPRKAFRRSSSEALCED 1100
QY 1157 AARPAGATLEBAKTLSPGKGVKVDVAFAGAVENPFYLPPOGGAADQHPHPAFAFD 1216
Db 1101 GAQCAGIFRVR-----FSGERN-----POGG----- 1122
QY 1217 NLVYMDQDPERGAPSTFKGTPTAENPE 1245
Db 1123 ----QQRKLTASSPSSFKTMADEDE 1146

```

RESULT 7
 Q9BIH9
 ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
 AC Q9BIH9;

```

DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_Taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/Genbank/DBD databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P1362; 1FGK.
DR InterPro; IPR000345; CycC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_th_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR PRINTS; PR01030; Recep_L_domain; 2.
DR PRODOM; PR00109; TYRKINASE.
DR PRODOM; PR000001; Euk_pkinase; 1.
DR SMART; SM00261; Fv; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.5%; Score 2012.5; DB 5; Length 1433;
Best Local Similarity 32.3%; Pred. No. 2.8e-143;
Matches 468; Conservative 202; Mismatches 390; Indels 387; Gaps 38;

QY 26 CTGTDMKRLRPASPEHNLDMRLHYGCGVVGNNLELYLFTFNASLSFLQDIOEVGYVL 85
Db 1 CIGTNGRMGVPANREHYNKLDRYNTCYVGNLEITWIQNTDNLNFQHIREVGYVL 60
QY 86 IAHNGVROYVRLRLRIVRGTOF-----EDNYALAVLDNGDPLNNTTPVTSAPGLREL 140
Db 61 ISYDLPQVILPRLOLIRKRTTFKLNKEEAYGLV-----SFSHMNTL 104
QY 141 QURSLTEILKGVLIQRLNOLCYODTILMKDI-FHKNQALATLIDTNRSRACHPSPKC 199
Db 105 ELPALRDLIGSGVGFNNVNLCHMSINMEEILAPQTSMTQTFNPSSEBVCPCHPSC 164
QY 200 KSGRCWGESSEDDCGSLTRVCAAGCA--RCKRPLPTDCCHCCAGCAAGCTGPKSDCLACH 257
Db 165 EVG-CWEGEAGHNCQRFSLKNCSPQSGCFGPKRPECCHLECAAGCTGTPQSDCLACH 223
QY 258 FHNSGICEJLHCPALVTYNTDTFESMPNPEGRTFPGASCYTAQYVNLSTDVSGCTLVCP 317
Db 224 FYDDGVCQEGCPFMQIYNPTNFWMEPNPBGKAYATGCRKP-EHLKDNACVYRKCPX 282
QY 318 HNOVTALEGTGRCCKSPCARVVCYGLAQWYIKANSKFIGITLEFAGCKKIFGSLAF 377
Db 283 GMPQNSE-----CVPCKGVCPKTCPEGEI-----VHSDNIG-----NYXDCITIEGSLLE 329
QY 378 PSFSGDPAASNT-----APLOEQLQVFELEITGYLYISAMPDSLPLSVFQNLQ 429
Db 330 DQSFQDFQGVYNTNFSFGPRYIKIDPDLRELVSTVAEITGFTNIQAHNRPFTLNTYFRNDE 389

```


QY 430 VIRGRIILANGAY-SLTLQGLIGISWLGRLSRELGSLALIHNTLHCFTVTPWDLFRN 468
 DB 390 VVGGRQKJENLPAVYIVTKSTLKLKSLKRVNVSISVILENSDLCEVEDIMSEIKKS 449
 QY 489 PHOALLTANPRECECEGEGELACHQOLARHGWPGTTCVNSQFLRGECECECVLQ 548
 DB 450 SDHEWVQKNNATECHEBMECEQCSKAGCKGKGFQCECNVYKXKCKLDSCK- 506
 QY 549 GLPREY-VNARHCLPCHPECCOPNGSVTCFGEADQCVACAHYKDPFCVARGP- 601
 DB 507 SLRPLYSVDSKTCGDCHQCKD-----FCGPNEDNGSCMNVDGRFCVACPTTGAM 561
 QY 602 -----SGVKPDLSTMPMKFPD----- 618
 DB 562 NGTCINCHKTCVGRGPRDTIAPDGCISCDKAIIGSADAKIERCLMKDESCPDGYSDYVL 621
 QY 619 -EEG----- 621
 DB 622 QEEGPKLQSGKAVCRKCHPRCKCTGTFHEQFCQECTGYKKGQCEDECPQDFYANEE 681
 QY 622 -ACQPCFINCT-----HSCVDL-----D-----KCGPABG----- 646
 DB 682 TRICLPCHQECRGCHGLDDHNERNLKLFEGDHYDVAATFTVCVSNCPASHPKRFPQEA 741
 QY 647 -----RASPLTSIVSAVVGILLVVLGVVFGI-LIFN--NFTVSGWL 685
 DB 742 GKGPFYCSADSMQSLRIEFQTKYMGVSMALILCVGFIAPVLFRRKCKKDAVNM 801
 QY 686 RVPKVASHLEPLTPSGAMPNQAQMRILKETELRKVYLSGAGFTYKGIWIDGENVK 745
 DB 802 TMAIAGCEDESEPLPSNVGPNLTKRIITKEAIEIRRGVLGAGAGRVFKGVWMEGESYK 861
 QY 746 IPAVAKVLRENTSPRANKELLIDBAVYVAGVSFVYSRLIGCLSTVQLVQLMPYCL 805
 DB 862 IPAVAKVMEWSGESSKSLBETATIMASVEHPNLIKLLAYCMISQKMLLIQLPGLCL 921
 QY 866 RLLDDIDETEHADGKVPKIMMALESILRRFTIOSDVSQVWELMTGAPYQIP 925
 DB 982 KLLPDSSEYRAAGKMPKIMWALECIRHRAVFTSKSDWAGITWELLTYGAPYEVNP 1041
 QY 986 AREIPDLLEKGERLPOPEICTIDVYIMVWCMMDSECRPRFRELVSFSMADPQRFV 985
 DB 1042 AKDVELIEIGHKLPQPDICSLDYCILSCVLDADARPTFOAETFAKADPGRYL 1101
 QY 986 VIQNEDDLGPASFLDSTFYRSLLEDMDKGLV----- 1016
 DB 1102 MI-----PGDKFMRLLPSYTNODEKDLIRTLAPVMAAAAAAAGASNVDPSTIA 1152
 QY 1017 DAEYLVPOQGFCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEEAAPS----- 1071
 DB 1153 EIDDEVLCQKTRISVLPGBSA-----VEPS-DMPKSLRYCK 1188
 QY 1072 -PLAP--SEGASGVFPDGLCMGAAGLQSLPHDPBPLORYSEDPVLPBSETGVYA 1127
 DB 1189 DPLKPDDETDGHEKVEV-----GVGIR-----LNPFLDEDDYLM 1222
 QY 1128 PLTCSPOPEYVQPVRRPQPSFREGPLPAARPAATLERAKTISPGKGVGVKVFAGG 1187
 DB 1223 P-TCQSQ-----NQS-----TPG-----YMDLIGIPA 1243
 QY 1188 AVENPEYV-----TPQGAAPQHPHPAPAFSAFNLVYWDQPPERGAAPSTFKGT 1238
 DB 1244 SYDNBEYLMGSTQAIAGAGSMG--PHTPP-----PNTPPGM 1280
 QY 1239 PTAENPE 1245
 DB 1281 PTHQHSQ 1287

RESULT 8
 Q9UK79 PRELIMINARY; PRT; 419 AA.
 ID Q9UK79;
 AC Q9UK79;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Heretatin.
 GN HER-2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99415951; PubMed=10485918;
 RA Doherty J.K., Bond C., Jardim A., Adelstein J.P., Clinton G.M.;
 RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
 RT autoinhibitor";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Doherty J.K., Clinton G.M., Adelstein J.P., Evans A.J., Henner W.D.;
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF177761; AAD5609.2; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 1.
 DR SMART: SM00261; FU; 1.
 SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;
 Query Match 27.5%; Score 1871; DB 4; Length 419;
 Best Local Similarity 98.8%; Pred. No. 2, 4e-133;
 Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MELALCRWGLLALLPQGAASVQCTGDMKRLPASBETHLDMRLHYOGQVQVNM 60
 DB 1 MELALCRWGLLALLPQGAASVQCTGDMKRLPASBETHLDMRLHYOGQVQVNM 60
 QY 61 ELYLPNMSLSLFDIOIGVQGVLLAHNVQVOPVQRIYRGQLFEDNYALAVLDNG 120
 DB 61 ELYLPNMSLSLFDIOIGVQGVLLAHNVQVOPVQRIYRGQLFEDNYALAVLDNG 120
 QY 121 DPLNTPVYVAGSPGRLRELQRLSLTEILKGVLLIQNPQLCYQDTILMKDIFPKXNOA 180
 DB 121 DPLNTPVYVAGSPGRLRELQRLSLTEILKGVLLIQNPQLCYQDTILMKDIFPKXNOA 180
 QY 181 LTLIDNRSRACHPCSPMCKSGRCWSESESDQSLRTYVAGGACACGKPLPTDCHEQC 240
 DB 181 LTLIDNRSRACHPCSPMCKSGRCWSESESDQSLRTYVAGGACACGKPLPTDCHEQC 240
 QY 241 AAGCTGKSHDCLACHFNHSGICELHCPALVYNTDFTESMNPGRYVFGASCYTACP 300
 DB 241 AAGCTGKSHDCLACHFNHSGICELHCPALVYNTDFTESMNPGRYVFGASCYTACP 300
 QY 301 YNLSIDVSGCLVCPILHNOEVTAEQTCQCEKCSRPCARVCGGL 345
 DB 301 YNLSIDVSGCLVCPILHNOEVTAEQTCQCEKCSRPCARVCGGL 345
 RESULT 9
 Q8R2X1 PRELIMINARY; PRT; 367 AA.
 ID Q8R2X1;
 AC Q8R2X1;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical 40.2 kDa protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC027080; AA027080.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;
 Query Match 25.5%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 2e-123;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;
 QY 887 MALESLIRRRFTQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 946
 DB 1 MALESLIRRRFTQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 60
 QY 947 IDVYIMVKCMWIDSECRPRFRELVSFSSRMARDPQRFVIVIONEDLGPSPLDSTYRSL 1006
 DB 61 IDVYIMVKCMWIDSECRPRFRELVSFSSRMARDPQRFVIVIONEDLGPSPLDSTYRSL 120
 QY 1007 LEDDDMGDLVAEEYLVQCGFFCPDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1066
 DB 121 LEDDDMGDLVAEEYLVQCGFFCPDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1067 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 1126
 DB 181 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 240
 QY 1127 APLTCSPOPEYVQPDVPPSPREGPLPAARPAATLERAKTLPKNGVVKDVFAG 1186
 DB 241 APLTCSPOPEYVQPDVPPSPREGPLPAARPAATLERAKTLPKNGVVKDVFAG 300
 QY 1187 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLYYNQDPERCAPSTKGTPTANPEY 1246
 DB 301 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLYYNQDPERCAPSTKGTPTANPEY 360
 QY 1247 LGADVPEV 1253
 DB 361 LGADVPEV 367
 RESULT 10
 Q8WYVO
 ID Q8WYVO PRELIMINARY; PRT; 412 AA.
 AC Q8WYVO:
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN PP3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF18349; AAL55856.1; -
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP motif.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;
 Query Match 24.9%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 3.3e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;
 QY 887 MALESLIRRRFTQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 946
 DB 1 MALESLIRRRFTQSDVMSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 60
 QY 947 IDVYIMVKCMWIDSECRPRFRELVSFSSRMARDPQRFVIVIONEDLGPSPLDSTYRSL 1006
 DB 61 IDVYIMVKCMWIDSECRPRFRELVSFSSRMARDPQRFVIVIONEDLGPSPLDSTYRSL 120
 QY 1007 LEDDDMGDLVAEEYLVQCGFFCPDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1066
 DB 121 LEDDDMGDLVAEEYLVQCGFFCPDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180
 QY 1067 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 1126
 DB 181 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 240
 QY 1127 APLTCSPOPEYVQPDVPPSPREGPLPAARPAATLERAKTLPKNGVVKDVFAG 1186
 DB 241 APLTCSPOPEYVQPDVPPSPREGPLPAARPAATLERAKTLPKNGVVKDVFAG 300
 QY 1187 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLYYNQDPERCAPSTKGTPTANPEY 1246
 DB 301 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLYYNQDPERCAPSTKGTPTANPEY 360
 QY 1219 YWMD-QDPPER-----GAPPSFKGTPTAN 1243
 DB 361 WWTQCEBEGVRRSPDVSSGREGTLTSGIKRWEKGPPTSGTCHARN 410
 RESULT 11
 Q86712
 ID Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 NT Polypeptide.
 UN POLYPEPTIDE.
 OS Avian fous-associated virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
 OC NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Venstrom B., Raynocheck C., Jansson L., Doederlein G., Uhocak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60725.1; -
 DR HSP: P03372; 1A6S
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR004028; Retro_M.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF02813; Retro_M; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 8402F6914EF61D63 CRC64;
 Query Match 24.1%; Score 1643; DB 15; Length 729;
 Best Local Similarity 53.8%; Pred. No. 1e-115;

Matches 344; Conservative 71; Mismatches 124; Indels 100; Gaps 14;

QY 582 DQVACAHYKDPFCVACPCSPGVRPDLSTYMPIMKFPDEGACQPCPCINCTHSCVDLDDKG 641
 DB 152 DHCMKCAHFIDGPHCVACPCAGVLGENDTL-VWKYADANAVCQLCHPCTRGCKGPGLEG 210
 QY 642 CPAEQRASPLTSIVSAV-GILLVVVLGVVFGILFNFTV-SFWLRVPKVSASHLEPLT 699
 DB 211 CP---NGSKTPSIAGVVGGLCLVVGGLGILYLRHRHVRKTRRLRLOERLEVEPLT 267
 QY 700 PSGAMPQAOQRILKETELRKVKYLSGAFGVYKGIIPDGENVKIPVAIKYLRENTSP 759
 DB 268 PSGEAPNQAHRLIKETEFKKVKYLSGAFGVYKGIIPDGENVKIPVAIKYLRENTSP 327
 QY 760 KANKELDEAVYVAGVSPYVSRLLGICITSTVQVLTQMPYGCILDVHRENGRGLGSD 819
 DB 328 KANKELDEAVYVAGVSPYVSRLLGICITSTVQVLTQMPYGCILDVHRENGRGLGSD 819
 QY 820 LNMCMQIAKGMSTYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADG 879
 DB 388 LNMCMQIAKGMSTYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADG 879
 QY 880 GXPVIRKMALESILRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERL 939
 DB 448 GXPVIRKMALESILRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERL 939
 QY 940 POPPCTIDVYIMVWKCMIDSECRPFRELVSFSEPMARDPQRFVYIO-NEDLGPASPL 998
 DB 508 POPPCTIDVYIMVWKCMIDSECRPFRELVSFSEPMARDPQRFVYIO-NEDLGPASPL 998
 QY 999 DSTFYRSLLEDDMGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 1058
 DB 568 DSTFYRSLLEDDMGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 1058
 QY 1059 LGLEPSEEARPSPL-----APSEAGSDVFPDGLGMAKGLQSLPTHPSPLQRYSED 1113
 DB 604 -----SRTPLLSSLSATSNNSATNCID-----RNGGCHPVRDEDSFYQRYSSD 645
 QY 1114 PTVPLEPSET--DGYVAPLTCSPOPEYVNOQDVRPQPSPREGRLPAARPAAGATLERAKTL 1171
 DB 646 PTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS-----T 676
 QY 1172 SPQKNGVVKDVF-----AFGAVENPEYL 1195
 DB 677 AMVQNOIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL 715

RESULT 12
 ID 086714 PRELIMINARY: PRT: 567 AA.
 AC 086714
 DT 01-NOV-1996 (TREMELREL 01, Created)
 DT 01-NOV-1996 (TREMELREL 01, Last sequence update)
 DT 01-1AR-2002 (TREMELREL 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-ERBB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.
 NC NCBI_Taxid=11950;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of C-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60727.1; .
 DR HSSB: P11362; 1FGK.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR001245; Tyr_Kinase.
 DR Pfam: PF00069; Pkinase; 1.

DR PRINTS; P800109; TYRKINASE.
 DR Prodom; P8000001; Euk_Pkinase; 1.
 DR SMART; SMO0219; Tyrc; 1.
 DR PROSITE; P500107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; P500101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; P800109; PROTEIN_KINASE_TYR; 1.
 KW Tyrosine-protein kinase.
 FT NON_TER
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 24.1%; Score 1641; DB 15; Length 567;
 Best Local Similarity 53.8%; Pred. 1e-115;
 Matches 344; Conservative 71; Mismatches 124; Indels 100; Gaps 14;

QY 582 DQVACAHYKDPFCVACPCSPGVRPDLSTYMPIMKFPDEGACQPCPCINCTHSCVDLDDKG 641
 DB 3 DHCMKCAHFIDGPHCVACPCAGVLGENDTL-VWKYADANAVCQLCHPCTRGCKGPGLEG 61
 QY 642 CPAEQRASPLTSIVSAV-GILLVVVLGVVFGILFNFTV-SFWLRVPKVSASHLEPLT 699
 DB 211 CP---NGSKTPSIAGVVGGLCLVVGGLGILYLRHRHVRKTRRLRLOERLEVEPLT 267
 QY 700 PSGAMPQAOQRILKETELRKVKYLSGAFGVYKGIIPDGENVKIPVAIKYLRENTSP 759
 DB 268 PSGEAPNQAHRLIKETEFKKVKYLSGAFGVYKGIIPDGENVKIPVAIKYLRENTSP 327
 QY 760 KANKELDEAVYVAGVSPYVSRLLGICITSTVQVLTQMPYGCILDVHRENGRGLGSD 819
 DB 328 KANKELDEAVYVAGVSPYVSRLLGICITSTVQVLTQMPYGCILDVHRENGRGLGSD 819
 QY 820 LNMCMQIAKGMSTYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADG 879
 DB 388 LNMCMQIAKGMSTYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADG 879
 QY 880 GXPVIRKMALESILRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERL 939
 DB 448 GXPVIRKMALESILRRFTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERL 939
 QY 940 POPPCTIDVYIMVWKCMIDSECRPFRELVSFSEPMARDPQRFVYIO-NEDLGPASPL 998
 DB 508 POPPCTIDVYIMVWKCMIDSECRPFRELVSFSEPMARDPQRFVYIO-NEDLGPASPL 998
 QY 999 DSTFYRSLLEDDMGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 1058
 DB 233 DSTFYRSLLEDDMGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 1058
 QY 1059 LGLEPSEEARPSPL-----APSEAGSDVFPDGLGMAKGLQSLPTHPSPLQRYSED 1113
 DB 299 LGLEPSEEARPSPL-----APSEAGSDVFPDGLGMAKGLQSLPTHPSPLQRYSED 1113
 QY 1114 PTVPLEPSET--DGYVAPLTCSPOPEYVNOQDVRPQPSPREGRLPAARPAAGATLERAKTL 1171
 DB 359 PTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS-----T 676
 QY 1172 SPQKNGVVKDVF-----AFGAVENPEYL 1195
 DB 528 AMVQNOIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL 566

RESULT 13
 ID 064895 PRELIMINARY: PRT: 962 AA.
 AC 064895
 DT 01-NOV-1996 (TREMELREL 01, Created)
 DT 01-NOV-1996 (TREMELREL 01, Last sequence update)
 DT 01-JUN-2002 (TREMELREL 21, Last annotation update)
 DE Gag-V-erbA-V-erbB protein.
 GN GAG-V-ERB-A-V-ERB-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
 NC NCBI_Taxid=11861;
 RN (1)
 RP SEQUENCE FROM N.A.

```

RX MEDLINE=90206603; PubMed=1969616;
RA Bruskini A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: X52209; CA36459.1; -.
DR EMBL: X52211; CA36459.1; JOINED.
DR HSSP: P10828; 2N1L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001723; Steroid_receptor.
DR InterPro: IPR001245; Tyr_pkinase.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec.1.
DR Pfam: PF00069; pkinase.1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STRDHOMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000001; Euk_pkinase.1.
DR ProDom: PD000035; Znf_C4steroid.1.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00219; Tyrc; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KM ATP-binding; DNA-binding; Nucleic acid binding; Tyrosine-protein kinase;
KM Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E495CE CRC64;

Query Match 23.1%; Score 1576.5; DB 15; Length 962;
Best Local Similarity 50.2%; Pred. No. 1,7e-110;
Matches 347; Conservative 71; Mismatches 154; Indels 119; Gaps 18;

QY 541 VEECRYLQGLPRE-YYNAR-HCLP-----CHPECO 568
DB 354 IEKCOESYLLAFHYIYRKHNIPEWSKLMKRVADLMIGAHASRFLMKVECPTELS 413
QY 569 POKGSVTCFGEPAQOCVACAHYKDPFCVACPSGVKPDSTYMPKPFDEBACQPCPI 628
DB 414 PGE-----VGP--DHCKKCAHFIDGPHCVACAGVIGENDTL-VKRYADANAVALCHP 465
QY 629 NCHTSCVDLDDKSCPAQASPLTSIVSAVV-GILLVVLGVVFGILIFNNFTV-SFWLR 686
DB 466 NCTRGCKGPGLEGP--NSKTPSIAAGVVGILLCLVVGILGIGYLRRIYKRTLR 522
QY 687 VPRVSAHLEPLTPSGAMPNOAMRIKETELRKVVLGSGAGTGYKGIWTPDGENVKI 746
DB 523 RLIGRELVEPLTPSGAPNOAHRLIKETEFKKVKVGLGAGFTYKGLM-PEGKVTI 582
QY 747 PVAIKVLENTSPRANKELIDEAYVYVAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLD 806
DB 583 PVAIKELREATSPRANKELIDEAYVYVAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLD 642
QY 807 HVENRGRGLSGQDLLWCMQIAGMSYLEBVLVHDLAARNLVKSPNNVKTIDPGLAR 866
DB 643 YIEHKDNIQSQYLLWMCVQIAGKMYLEERHVAHDLAARNLVKTPHVKITDGLAK 702
QY 867 LLDIDETEVHADGKVPKIMMALESLIRRFTHQSDVWSYGVTWELMTFGAPYDGI 926
DB 703 QLGADKEVYHAGGVKVPKIMMALESLIRRFTHQSDVWSYGVTWELMTFGAPYDGI 762
QY 927 RELPDLLEKGERLPQPPICITDVTYVIMVKCMIDSECRPFRELVSEFRMAADPORFV 986
DB 763 SEISSVLEKGERLPQPPICITDVTYVIMVKCMIDSECRPFRELVSEFRMAADPORFV 822
QY 987 IQ-NEDLGPASPLDSTFYRSLLEDDMGDLVDAEYLVLPQGGFCDDPAPAGAGMWHHH 1045
DB 823 IQGDERKHLPSPTDSKFTYRSLLEDDMGDLVDAEYLVLPQGGFCDDPAPAGAGMWHHH 866

```

```

QY 1046 RSSSTRSGGDLTLLEPSEBEPAPRPLAPSRGAGSDVPDGLGMAKGLQSLPTHDS 1105
DB 867 NSPST-----SKTPLLSSLSATSN-----NSATKICIRNGCH--- 898
QY 1106 PLQRYSEDPTEVLPSEIDGVYAPLTCSPQPEYVNOPDVPQPPSPREGPLPAPAPAGAT- 1164
DB 899 -----PREDGFL-----PAPEYVNG--LMPKKFSTAMQNOQYNYISLTA 937
QY 1165 LERATLSPKXGVYKDVFAFGAVENPEYL 1195
DB 938 ISKLPMDSRYQN-----SHSTAVDNEPYL 961

RESULT 14
ID 085468 PRELIMINARY; PRT; 545 AA.
AC 085468;
DT 01-NOV-1996 (TREMBLER). 01, Created)
DT 01-NOV-1996 (TREMBLER). 01, Last sequence update)
DT 01-MAR-2002 (TREMBLER). 20, Last annotation update)
DE Avian erythroblastosis virus (T334) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_Taxid=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scolding P., Venstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL: X06943; CA30024.1; -.
DR HSSP: P11362; 1FCR.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase.1.
DR ProDom: PD000001; Euk_pkinase.1.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KM ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DC88CA0FAPFA CRC64;

Query Match 23.0%; Score 1568; DB 15; Length 545;
Best Local Similarity 53.3%; Pred. No. 3,2e-110;
Matches 334; Conservative 68; Mismatches 133; Indels 92; Gaps 15;

QY 578 GREADQCVACAHYKDPFCVACPSGVKPDSTYMPKPFDEBACQPCPICTHSCVDL 637
DB 1 GP--DHCKKCAHFIDGPHCVACAGVIGENDTL-VKRYADANAVALCHPCTGCKGP 57
QY 638 DKGCPAQRASPLTSIVSAVV-GILLVVLGVVFGILIFNNFTV-SFWLRPKVSASHL 695
DB 58 GLEGP--NSKTPSIAAGVVGILLCLVVGILGIGYLRRIYKRTLRLLDRELY 114
QY 696 EPLTPSGAMPNOAMRIKETELRKVVLGSGAGTGYKGIWTPDGENVKIPVAIKVLE 755
DB 115 EPLTPSGAPNOAHRLIKETEFKKVKVGLGAGFTYKGLM-PEGKVTI 702
QY 756 NTPSPRANKELIDEAYVYVAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLD 806
DB 115 ATSPRANKELIDEAYVYVAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLD 762
QY 807 HVENRGRGLSGQDLLWCMQIAGMSYLEBVLVHDLAARNLVKSPNNVKTIDPGLAR 866
DB 703 QLGADKEVYHAGGVKVPKIMMALESLIRRFTHQSDVWSYGVTWELMTFGAPYDGI 762
QY 867 LLDIDETEVHADGKVPKIMMALESLIRRFTHQSDVWSYGVTWELMTFGAPYDGI 926
DB 703 QLGADKEVYHAGGVKVPKIMMALESLIRRFTHQSDVWSYGVTWELMTFGAPYDGI 762
QY 927 RELPDLLEKGERLPQPPICITDVTYVIMVKCMIDSECRPFRELVSEFRMAADPORFV 986
DB 763 SEISSVLEKGERLPQPPICITDVTYVIMVKCMIDSECRPFRELVSEFRMAADPORFV 822
QY 987 IQ-NEDLGPASPLDSTFYRSLLEDDMGDLVDAEYLVLPQGGFCDDPAPAGAGMWHHH 1045
DB 823 IQGDERKHLPSPTDSKFTYRSLLEDDMGDLVDAEYLVLPQGGFCDDPAPAGAGMWHHH 866

```

QY 936 GERLPDPICITIDVYIMVCMWIDSECRPRELVESEFMAADPREFVIO-NEDLCP 994
 DB 355 GERLPDPICITIDVYIMVCMWIDSECRPRELVESEFMAADPREFVIO-NEDLCP 414
 QY 995 ASPIDSTFVSLLEDDMDGLVDAEFLVQOQFCDDPAPGAGMWHHRSSSRSGC 1054
 DB 415 PSPIDSKFRTLMSEEDMEDVDAEFLVPHQGF-----NSPST----- 454
 QY 1055 GDLTLGLPESEEARPSPL-----APSEGASDVFDGLGMAKGLQSLPTHDPSPLO 1109
 DB 455 -----SRTPLLSLSATSNMSTNCIDRNGC-----H----- 481
 QY 1110 YSEDPYPLPSETDGYAPLTCSPQPEYVNOQPPQPPSPREGPLPARPAGT-LE 1168
 DB 482 -----PVREDFPL-----PAPRYVNO-LMPKPPSTAMVQIYVYISLTLAKL 524
 QY 1169 KTLSPGKGVVGVDFARFAGAVENPEYL 1195
 DB 525 PMDSRYON-----SHSTAVDNPEYL 544

RESULT 15

Q9WVF5

PRELIMINARY:

PRT: 655 AA.

AC Q9WVF5:

DT 01-NOV-1999 (Tremblrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).

GN EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090.

RN (1)

RN SEQUENCE FROM N.A.

RN RC STRAIN=C57BL/6J; TISSUE=LIVER;

RN RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,

RN RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RN RA Maibhe N.J.;

RT "Alternative Transcripts from the Human and Mouse EGFR Gene Encode

RT Carboxy-Terminal Truncated Receptors";

RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.

RN (2)

RN SEQUENCE FROM N.A.

RN RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAJ;

RN RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,

RN RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,

RN RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,

RN RA Maibhe N.J.;

RT "Comparative genomic sequence analysis and isolation of human and

RT mouse alternative Egfr transcripts encoding truncated receptor

RT isoforms";

RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

RN (3)

RN SEQUENCE FROM N.A.

RN RC STRAIN=C57BL/6J; TISSUE=LIVER;

RN RX MEDLINE=21085660; PubMed=11217851;

RN RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RN RA Aizawa K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,

RN RA Saito T., Okazaki Y., Gojobori T., Bono H., Kaikawa T., Saito R.,

RN RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RN RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,

RN RA Kuehl P., Lewis S., Maeno Y., Nikaide I., Pesole G., Quackenbush J.,

RN RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RN RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RN RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RN RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RN RA Guernstein S., Hill D., Hofmann M., Hume D.A., Kamitani M., Lee N.H.,

RN RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombereis P.,

RN RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RN RA

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection";

RL Nature 409:685-690 (2001).

DR EMBL: AF124513; RAD44149.1; -

DR EMBL: AF275366; AAG28047.1; -

DR EMBL: AF275364; AAG28047.1; JOINED.

DR EMBL: AF275365; AAG28047.1; JOINED.

DR EMBL: AK004944; BAB23688.1; -

DR EMBL: AK004883; BAB23641.1; -

DR EMBL: AK004911; BAB23662.1; -

DR MGD: MG1:95294; Egfr.

DR InterPro: IPR000494; EGFR_L_domain.

DR InterPro: IPR002174; Furin-like.

DR Pfam: PF00757; Furin-like; 1.

DR Pfam: PF01030; Recep_L_domain; 2.

DR SMART: SM00261; FU; 3.

KM Receptor.

SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.1%; Score 1506.5; DB 11; Length 655;

Best Local Similarity 44.5%; Pred. No. 1.9e-105;

Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 L1LALLPGAA--STQVCTGTDMKRLPASPETHLDMRLHYOGGVVQGNLELYPTN 68

DB 14 L1LALCAAGALEBKVKVCGTNRNLQGLTFEDHFLSLDRMTNNEVVLNLEIITYQGN 73

QY 69 ASLSFLQDIQEVGYVLLAHNOVROVPLRLRIVEGTOLFEENVALAVLDNGDPLNNTTP 128

DB 74 YDLSPFLKTIQEVAGVLLALNTERPLENTLQIIGNALYENTVALALISN----- 124

QY 129 VTGASPGGLRELQSLSTELIKGVYLIGNPQLQVDTLMDI----FKKNQALATLI 184

DB 125 -YGNRTGRLREPLMRQLQETLLGVAFSNNPLLCNMDITQMRDIVQNVFMSMSMDL- 180

QY 185 DTNRSPACHPGSPMCKGRCMWESSEDCSLTRTYCAGCA-RCKGPLPTDCBQCAAG 243

DB 181 -QSHPESSCKDPSCKNGSCWGGGEECKLTKIICAOQSHRCKGRSPDCCHQCAAG 239

QY 244 CTGPKASDLCLAFHNSGICELHCAVLTNTDTFESMPNPEGKTFPASCVTAPRY 303

DB 240 CTGPRSDCLVQCKPDEATCKDTGPEPLMLNPTTYQMDVNEGKSFATVCKCPRY 299

QY 304 LSTDVSCCTVCPPLHNOEYTAEDGTORCEKSCPCARVCYGLGWQYIKANSKFGITEL 363

DB 300 VVTDHSCYRACQPDYEV-EDDGIRKCKCKGCPKRYNGIGIGEFK-DTISINATNK 357

QY 364 -FAGCKKIFGSLAFPESPDGPASNTAFLOEQLOVFTLEITGYLYISAMPDLP 422

DB 358 HFKYCTAIGDHLIIFVAFKGSFETPLPRELEILTYVEIGFLLIQAMPDMWT 417

QY 423 SVFQNLQVIRGRILHNGAVSLTQGGIWMLGIRSLREKSGGALILHNHTLCFVATPM 482

DB 418 HAFENLEIRGRKQHGQFSLAVGLNTSLGRSLKEISDGVYIISGRNLQYNTIM 477

QY 483 DQLFNPHQALLTANRPDECEVGEGLAQHQLCARGCMGPGPTQCVNCSQFLQOEVE 542

DB 478 KKLFGTPNQKTKIMNRARAKDKAVHNVCPNCSGSGCPEPRDVCVSNVSRGECVE 537

QY 543 ECVVLQGLPREVYNAHCLPCHPEQOPQMSGVTCEGPEADOCVACHYDPPEFCVARS 602

DB 538 KCVLTGEREFREVENSECTQCHBELCPQMANITCTGRGDNCTQCAHYDGHCVCTGA 597

QY 603 GYVPSLSTYPIWKPDEBAGQCPPICTNCTHSCVLDLQK 642

DB 598 GINGENNTL-VWKYADANNVCHLCHANCYGCAGPQLQCC 636

Search completed: July 22, 2003, 09:24:29

Job time: 51.75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      July 22, 2003, 08:41:54 ; Search time 36.875 Seconds
              (without alignments)
              4527.811 Million cell updates/sec
```

Title: SEQ4-653-675-14

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1253

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

```
Database :
A.Genseq.101002.*
1: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/genseq/genseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/genseq/genseqp-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	6610	97.0	1255	21	AA9Y62620	Human hercyeulin 2
2	6610	97.0	1255	22	AA612130	Human tyrosine kinase
3	6610	97.0	1255	22	AA660167	HER2 transgene plasmid
4	6610	97.0	1255	23	AA674545	Human HER2 (ErbB2)
5	6604	96.9	1255	17	AAW01111	Human HER2/neu protein
6	6604	96.9	1255	20	AAW92406	Human HER-2/neu protein
7	6604	96.9	1255	21	AA621198	Human HER-2/neu protein
8	6604	96.9	1255	21	AA647480	Amino acid sequence of human HER-2/neu protein
9	6604	96.9	1255	22	AA854558	Human HER-2/neu protein
10	6604	96.9	1255	22	AA688267	HER2/neu amino acid sequence

45	2977	43.7	587	23	AAE20481	Human protein for
44	2977	43.7	583	23	AAE20483	Human protein for
43	3028	44.4	1210	23	ABP51768	Human epidermal gr
42	3053.5	44.8	654	21	AAE51150	Rat Her-2/neu onco
41	3053.5	44.8	654	21	AAE51125	Rat Her-2/neu prot
40	3067	45.0	1210	22	AAE68470	Amino acid sequen
39	3069	45.0	1210	23	AAE50768	Human epidermal g
38	3069	45.0	1210	21	AAE2301.9	Human Her-1 protei
37	3069	45.0	1210	21	AAE50616	Human EGF receptor
36	3069	45.0	1210	21	AAE19259	Amino acid sequen
35	3344	49.1	951	21	AAE08222	Extracellular portr
34	3447	50.6	951	21	AAE44993	Extracellular portr
33	3512	51.5	645	22	AAE61593	Human ErbB2 extrac
32	3512	51.5	645	22	AAE60408	Human ErbB2 oncop
31	3550	52.1	653	21	AAE51145	Human Her-2/neu o
30	3550	52.1	653	21	AAE21200	Human Her-2/neu o
29	3552	52.1	782	18	AAE19764	Human Her-2/neu f
28	3699	54.3	712	21	AAE51149	Her-2/neu extracel
27	3699	54.3	712	21	AAE21204	Human HER-2/neu f
26	4065.5	59.7	926	23	AAE51153	Mouse Her-2/neu ex
25	4065.5	59.7	926	23	AAE51152	Mouse Her-2/neu ex
24	4815	70.7	919	21	AAE51148	Human HER-2/neu f
23	4815	70.7	919	21	AAE511203	Human Her-2/neu f
22	5794.5	85.0	1256	22	AAE51151	Mouse Her-2/neu o
21	5794.5	85.0	1256	21	AAE21206	Mouse Her-2/neu p
20	5794.5	85.0	1256	21	AAE21206	Mouse Her-2/neu p
19	5821.5	85.4	1256	23	AAE51144	Rat Her-2/neu onco
18	5821.5	85.4	1256	21	AAE51144	Rat Her-2/neu onco
17	6287	92.3	1200	21	AAE21208	Human Her-2/neu p
16	6440	94.5	1223	23	AAU98923	Human Breast canc
15	6561	96.3	1433	14	AAE39568	Sequence of c-erbB
14	6604	96.9	1255	23	AAU77114	Human Her-2/neu p
13	6604	96.9	1255	23	AAE51143	Human Her-2/neu o
12	6604	96.9	1255	23	AAE20479	Human Her-2/neu p
11	6604	96.9	1255	23	AAE20479	Human Her-2/neu p
10	6604	96.9	1255	23	AAE20479	Human Her-2/neu p

ALIGNMENTS

RESULT 1	ID	AA92620 standard; Protein; 1255 AA.
XX	AA92620;	
AC	10-AUG-2000 (first entry)	
XX		
DT		
XX		
DE	Human heregulin 2 (Her2).	
XX		
KM	Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity	
KM	self-protein; cancer; breast cancer; prostate cancer;	
KM	cell-associated peptide antigen; foreign epitope.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Domain	1..173
FT		/label= "N-terminal
FT		/note= "mature polypeptide"
FT	Region	5..25
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT		/label= "insertion region
FT		/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT /label= Cysteine_rich_domain
 FT Region
 FT 210..224
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 250..264
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 369..383
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 465..479
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 653..667
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 695..709
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 710..730
 FT /label= insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 1011..1235
 FT /label= C-terminal_domain

WO200020027-A2.
 13-APR-2000.
 05-OCT-1999; .99WO-DK00525.
 05-OCT-1998; 98DK-0001261.
 20-OCT-1998; 98US-0105011.
 (MEBI-) M & E BIOTECH AS.
 Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,
 Gautam A, Birk P, Karlsson G;
 WPI: 2000-349917/30.
 N-PSDB; AAA09455.
 Inducing immune responses to weakly immunogenic, tumor associated
 peptide antigens for the treatment of breast and prostate cancer
 Claim 62; Page 193-198; 220pp; English.

This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 Her2 can be used in the claimed method as an autovaccine to induce a CTL
 response. Subdominant CTL epitopes, antibody binding regions and
 cysteine residues involved in disulfide bonds are preserved in the
 immunogenized forms. Regions suitable for the insertion of foreign T
 helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA, and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/preast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.

XX Sequence 1255 AA:

Query Match 97.0%; Score 6610; DB 21; Length 1255;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;

QY 1 MEIAAICRMGLLALLPPGAATQVCTGDMKRLPASPEHLLDMLRHLYQCGVVOGNL 60
 DB 1 MEIAAICRMGLLALLPPGAATQVCTGDMKRLPASPEHLLDMLRHLYQCGVVOGNL 60
 QY 61 ELTYLPNNASTFLQDIQEVQGVVLAHNOVROVPLORIRIVAGTOLFEDNVALAVLDNG 120
 DB 61 ELTYLPNNASTFLQDIQEVQGVVLAHNOVROVPLORIRIVAGTOLFEDNVALAVLDNG 120
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLLIQRNPLQCYQDTLLMKDIFHKNNOLA 180
 DB 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLLIQRNPLQCYQDTLLMKDIFHKNNOLA 180
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLLIQRNPLQCYQDTLLMKDIFHKNNOLA 180
 DB 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLLIQRNPLQCYQDTLLMKDIFHKNNOLA 180
 QY 181 LTLIDPTRSRAPCHPCSMCKSGRCWSESSDCSLRTVAGGACARCKGLPDDCCEQC 240
 DB 181 LTLIDPTRSRAPCHPCSMCKSGRCWSESSDCSLRTVAGGACARCKGLPDDCCEQC 240
 QY 241 AAGCTGPKHSDCLAFHFNHSGICELHCPALVTYNTDTPESMNPPEGRTYFGASCTACP 300
 DB 241 AAGCTGPKHSDCLAFHFNHSGICELHCPALVTYNTDTPESMNPPEGRTYFGASCTACP 300
 QY 301 YNYISTDVSGCTLVCPAHNOEYVAEDGTQCEKSCRPACVACGLGMQYIKANSKFTGIT 360
 DB 301 YNYISTDVSGCTLVCPAHNOEYVAEDGTQCEKSCRPACVACGLGMQYIKANSKFTGIT 360
 QY 361 ELBPAQCKKIFGSLAFPESEFDDPASNTPALQPEQLVETLEITGYLYISAMPDPLP 420
 DB 361 IQBPAQCKKIFGSLAFPESEFDDPASNTPALQPEQLVETLEITGYLYISAMPDPLP 420
 QY 421 DLSVFQNLQVIRGILHNGAYSLTLOGLSISGLSLRELSGLALIHNTHLCPVHTV 480
 DB 421 DLSVFQNLQVIRGILHNGAYSLTLOGLSISGLSLRELSGLALIHNTHLCPVHTV 480
 QY 481 PMDQLFRNPQALHTANRPEDCEVGGALCHQLCARGHQMGPGPTQCVNCSOFLRQEC 540
 DB 481 PMDQLFRNPQALHTANRPEDCEVGGALCHQLCARGHQMGPGPTQCVNCSOFLRQEC 540
 QY 541 VEECRVLOGLPREVYNAHCLPCHPECPONGSVTCGPBADCCVACAHYKDPFCVARC 600
 DB 541 VEECRVLOGLPREVYNAHCLPCHPECPONGSVTCGPBADCCVACAHYKDPFCVARC 600
 QY 601 PSGVPRDLSYMPIMKPFDEEGACOPCINCTHSCVDLDDGCAEORASPLTISAVVG 660
 DB 601 PSGVPRDLSYMPIMKPFDEEGACOPCINCTHSCVDLDDGCAEORASPLTISAVVG 660
 QY 661 ILVVVLGVVFGILV-----FNNFTVSWLRVKYSASHLEPLTSGAMPNOAKRILK 714
 DB 661 ILVVVLGVVFGILV-----FNNFTVSWLRVKYSASHLEPLTSGAMPNOAKRILK 714
 QY 715 ETELRKRVUGSAGFGVYGVGIMIPDENVKIYVALVLENTSPRANKELDEAYVMAG 774
 DB 715 ETELRKRVUGSAGFGVYGVGIMIPDENVKIYVALVLENTSPRANKELDEAYVMAG 774

QY 775 VGSPPVSRLLGICLTSTVQLVLTQIMPYGCLLDHVRNRRGLSGDILLNMCQAKGMSYL 834
 PS 777 VGSPPVSRLLGICLTSTVQLVLTQIMPYGCLLDHVRNRRGLSGDILLNMCQAKGMSYL 836
 Db 835 EDRLVHRDLAARNVLKSPNHYKITDFGLARLLIDETRYHADGKVPKXMALESILR 894
 Db 837 EDRLVHRDLAARNVLKSPNHYKITDFGLARLLIDETRYHADGKVPKXMALESILR 896
 QY 895 RRTFHQSDVWSYGTWELMTFGAKPYDGIIPAREIDPLEKGERLPPPICTIDVYMINV 954
 Db 897 RRTFHQSDVWSYGTWELMTFGAKPYDGIIPAREIDPLEKGERLPPPICTIDVYMINV 956
 QY 955 KCMWIDSECRPRRELVSERSMARDPQRFVVTQNEBGLASPJLSTFYRSLIEDDDMGD 1014
 Db 957 KCMWIDSECRPRRELVSERSMARDPQRFVVTQNEBGLASPJLSTFYRSLIEDDDMGD 1016
 QY 1015 LVDAEEYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDDTLGLPSEEBAPRSPILA 1074
 Db 1017 LVDAEEYLVPOQGFCDPAPGAGMWHRRSSSTRSGGDDTLGLPSEEBAPRSPILA 1076
 QY 1075 PSEGAGSDVDFDGLGMAAGLQSLPTHDSPLOQRYSEDETVLPSSETDGYVAPLTCSPQ 1134
 Db 1077 PSEGAGSDVDFDGLGMAAGLQSLPTHDSPLOQRYSEDETVLPSSETDGYVAPLTCSPQ 1136
 QY 1135 PEYVNOPDVRPOPSPSPSEGBLPAPAPGATLERAKTILSPKNGVWVDVAFGAVENPEY 1194
 Db 1137 PEYVNOPDVRPOPSPSPSEGBLPAPAPGATLERAKTILSPKNGVWVDVAFGAVENPEY 1196
 QY 1195 LTFQGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYGLDVPV 1253
 Db 1197 LTFQGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2
 ID AAE12130 standard; Protein: 1255 AA.
 XX AAE12130:
 AC AAE12130:
 DT 18-DEC-2001 (first entry)
 XX
 DE Human tyrosine kinase-type receptor, HER-2.
 XX
 KW Therapeutic compound; major histocompatibility complex; vaccine;
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 KW antigen presenting cell; human; tyrosine kinase-type receptor.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 774..782
 FT /note="Antigenic epitope"
 XX
 PN WO200166677-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 16-MAR-2001; 2001MO-US40328.
 XX
 PR 16-MAR-2000; 2000US-0527487.
 XX
 PA (GENZ) GENZYME CORP.
 XX
 PI Nicolette CA;
 XX
 DR WPI: 2001-616284/71.
 DR N-PSDB; AAD19731.
 XX
 PT Novel synthetic therapeutic compound for inducing immune response and
 PT for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX
 PS Claim 4; Page 63-67; 69pp; English.
 XX
 CC The invention relates to synthetic therapeutic compounds (antigenic
 CC peptides) with enhanced binding to major histocompatibility complex
 CC (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterized by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of gene
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 CC (774-782).
 CC
 SQ Sequence 1255 AA;
 Query Match 97.0%; Score 6610; DB 22; Length 1255;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;
 QY 1 MELAALCRWGLLALLPQGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVYQNL 60
 Db 1 MELAALCRWGLLALLPQGAASVQCTGDMKRLPASPETHLDMRLHYOGCQVYQNL 60
 QY 61 ELTYPTNASLFLDDIOEVGYVLAHNOYROYLQRLRYRGQLFEDNYALAVLNG 120
 Db 61 ELTYPTNASLFLDDIOEVGYVLAHNOYROYLQRLRYRGQLFEDNYALAVLNG 120
 QY 121 DELNNTPTVGTASPGELRELORSYTEILKGVLIQRNPOLCYODTILMKDIFHNKOLA 180
 Db 121 DELNNTPTVGTASPGELRELORSYTEILKGVLIQRNPOLCYODTILMKDIFHNKOLA 180
 QY 181 LTLIDTNSRACHPSPCKSGRCWGESSEDCQSLTRVCAAGCARCGPLPTDCCHQC 240
 Db 181 LTLIDTNSRACHPSPCKSGRCWGESSEDCQSLTRVCAAGCARCGPLPTDCCHQC 240
 QY 241 AAGCTGPKHSDCLALHFNHSGICEHCPALVYVTDTFESMPNDEGRYTGASCVTACP 300
 Db 241 AAGCTGPKHSDCLALHFNHSGICEHCPALVYVTDTFESMPNDEGRYTGASCVTACP 300
 QY 301 YNYLSTDVSGCTVCPHNOEYTAEDGTQRCCKSKPCARVCYGLMGYIRANSKFIGIT 360
 Db 301 YNYLSTDVSGCTVCPHNOEYTAEDGTQRCCKSKPCARVCYGLMGYIRANSKFIGIT 360
 QY 361 ELFPAGCKKRTGSLAFLESPFDGPASNTAPLOPQLOVFTLEITGVYISAMPDLP 420
 Db 361 IOEPAGCKKRTGSLAFLESPFDGPASNTAPLOPQLOVFTLEITGVYISAMPDLP 420
 QY 421 DLVSFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRLRELSGALILHNHTHLCFVATV 480
 Db 421 DLVSFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRLRELSGALILHNHTHLCFVATV 480
 QY 481 PWDQLEFRNPQALLHTARPEDECVGBGLACHOLCARHCGPPTQCVNCSQFLRGQEC 540
 Db 481 PWDQLEFRNPQALLHTARPEDECVGBGLACHOLCARHCGPPTQCVNCSQFLRGQEC 540
 QY 541 VEECGVLTQGLPREYVNAHCLPCHPECCOPONGSVTCFPEADOCVACHYADPPCVARC 600
 Db 541 VEECGVLTQGLPREYVNAHCLPCHPECCOPONGSVTCFPEADOCVACHYADPPCVARC 600
 QY 601 PSGVAPDLSYMPIMWTFPEBEGACOPCINCHSCVDLDDKCGPAEQRASPLTISYSAVVG 660
 Db 601 PSGVAPDLSYMPIMWTFPEBEGACOPCINCHSCVDLDDKCGPAEQRASPLTISYSAVVG 660
 QY 661 ILVVVLGVVFGIIL-----FNNFTVAFMVRVPVSASHLEPLTPSGAMPNOQMRILX 714
 Db 661 ILVVVLGVVFGIILKRNQKIRKTYMRLLOETEL---VEPLTPSGAMPNOQMRILX 716

```

QY 715 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVRENTSPKANKETLDEAYVMAG 774
DB 717 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVRENTSPKANKETLDEAYVMAG 776
QY 775 VGSPPVSRLLGICLTSTVQLVTOQMPYGCGLDHRRENRGRSGSODLNMCMQIAKMSYL 834
DB 777 VGSPPVSRLLGICLTSTVQLVTOQMPYGCGLDHRRENRGRSGSODLNMCMQIAKMSYL 836
QY 835 EDVRLVHRDLAARNVLVKSPNHVKITDPGLARLLDIDETEHADGKVPKIMALESILR 894
DB 837 EDVRLVHRDLAARNVLVKSPNHVKITDPGLARLLDIDETEHADGKVPKIMALESILR 896
QY 895 RRFTHQSDVMSYGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 954
DB 897 RRFTHQSDVMSYGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 956
QY 955 KCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQNEDIGPASPIDSTFYRSLLEDDMGD 1014
DB 957 KCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQNEDIGPASPIDSTFYRSLLEDDMGD 1016
QY 1015 LVDAEYLVPOGGFPCDPAPAGAMVHHRSSSTRSGSGLTLGLEPSEEAPRSPLA 1074
DB 1017 LVDAEYLVPOGGFPCDPAPAGAMVHHRSSSTRSGSGLTLGLEPSEEAPRSPLA 1076
QY 1075 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOYSEDPVPLPSETDGYVAPLTCSPQ 1134
DB 1077 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOYSEDPVPLPSETDGYVAPLTCSPQ 1136
QY 1135 PEYVNPDPVRPQPSPREGLPAPAPAGATLEAKTISPGKNGVYKVPFAFGAVENPEY 1194
DB 1137 PEYVNPDPVRPQPSPREGLPAPAPAGATLEAKTISPGKNGVYKVPFAFGAVENPEY 1196
QY 1195 LTPOGGAAPQHPPEPAPFADNLVYWDQDPPERGAPSPSTFKGTPTAENPEYGLADVPV 1253
DB 1197 LTPOGGAAPQHPPEPAPFADNLVYWDQDPPERGAPSPSTFKGTPTAENPEYGLADVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
DT 03-APR-2001 (first entry)
XX
XX
XX HER2 transgene plasmid construct encoded protein.
XX
XX Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX MO200100244-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-US172229.
XX
XX 25-JUN-1999; 99US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
XX
XX Erickson S, Schwall R;
XX
XX WPI: 2001-061962/07.
XX N-P3DB: AAF24297.
XX
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
XX receptor and does not respond to an anti-ErbB antibody, comprises
XX conjugating the antibody to a maytansinoid -

```

```

PS Example 3; Fig 4; 92pp; English.
XX
XX The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
XX Sequence 1255 AA;
SQ
Query Match 97.0%; Score 6610; DB 22; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;

QY 1 METALALCRMGILLALIPGAASVQCTGDMKLRPASBETHLDMRLHYQCGVQGNL 60
DB 1 METALALCRMGILLALIPGAASVQCTGDMKLRPASBETHLDMRLHYQCGVQGNL 60
QY 61 ELTYLPFNASLSTFDIOIEVQGVYLIANQVAVPQRIYRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPFNASLSTFDIOIEVQGVYLIANQVAVPQRIYRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLRELDLSLTELKGGVLIQRNPQLCYODTILMKDIFHXNQLA 180
DB 121 DPLNNTPTVYGASPGGLRELDLSLTELKGGVLIQRNPQLCYODTILMKDIFHXNQLA 180
QY 181 LTLIDPNRBRACHPGSPMKCKSGRCSWSESDOSLRTVCAGGACARCKPLPTDCHEOC 240
DB 181 LTLIDPNRBRACHPGSPMKCKSGRCSWSESDOSLRTVCAGGACARCKPLPTDCHEOC 240
QY 241 AAGCTGPKXSDCLACLFHFNHSGICELHCALVTYNTDRESMNPREGRTVFGASCYACP 300
DB 241 AAGCTGPKXSDCLACLFHFNHSGICELHCALVTYNTDRESMNPREGRTVFGASCYACP 300
QY 301 YNLTSTDVSGCTLVCLHNOEVTABDGTORCEKSPCARVCYGLCMQYIKANSKIGIT 360
DB 301 YNLTSTDVSGCTLVCLHNOEVTABDGTORCEKSPCARVCYGLCMQYIKANSKIGIT 360
QY 361 ELIEFACCKIFGSLAFIPESFDGDPASNTAPQPEOLQVETLEETGLYISAMPDILP 420
DB 361 IQEPACCKIFGSLAFIPESFDGDPASNTAPQPEOLQVETLEETGLYISAMPDILP 420
QY 421 DLSVFQNLQVIRGRIIHNAYSLTLQGLISWLGRLSRLBELSGALLIHNHNLCHVHTV 480
DB 421 DLSVFQNLQVIRGRIIHNAYSLTLQGLISWLGRLSRLBELSGALLIHNHNLCHVHTV 480
QY 481 PMDQLFRNPHQALHTANRPEDECVSEGLACHQLCARGHGWGPGPQCVNCSOPLRGQEC 540
DB 481 PMDQLFRNPHQALHTANRPEDECVSEGLACHQLCARGHGWGPGPQCVNCSOPLRGQEC 540
QY 541 VEECRVLYQLPREYVNAHCLPCHPEQCPQNSVITCFGEADQCVACAHYKDPFCVARG 600
DB 541 VEECRVLYQLPREYVNAHCLPCHPEQCPQNSVITCFGEADQCVACAHYKDPFCVARG 600
QY 601 PSGVRPDLSTYMPITWKPDEEGACQPCPINCTSVDLDKKGPAERAPLTSIYAVNG 660
DB 601 PSGVRPDLSTYMPITWKPDEEGACQPCPINCTSVDLDKKGPAERAPLTSIYAVNG 660
QY 661 ILLVVLGVVFGILLI-----FNNFTVSFMLEVPKXASASLPLTPSGAMPNQAQMRILK 714
DB 661 ILLVVLGVVFGILLIIRROOKIRKTYMRLLQETL-----VEPLTPSGAMPNQAQMRILK 716
QY 715 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVRENTSPKANKETLDEAYVMAG 774
DB 717 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVRENTSPKANKETLDEAYVMAG 776
QY 775 VGSPPVSRLLGICLTSTVQLVTOQMPYGCGLDHRRENRGRSGSODLNMCMQIAKMSYL 834
DB 777 VGSPPVSRLLGICLTSTVQLVTOQMPYGCGLDHRRENRGRSGSODLNMCMQIAKMSYL 836
QY 835 EDVRLVHRDLAARNVLVKSPNHVKITDPGLARLLDIDETEHADGKVPKIMALESILR 894

```

Db 837 EDVRLVHRLAARNVLVKSPNHVKTDFGLARLDIDETEVHADGKVPKXMALESILR 896
 QY 895 RRETHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLTKGRLPPICITIDYXIMW 954
 Db 897 RRTTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLTKGRLPPICITIDYXIMW 956
 QY 955 KCMWIDSECRPREFELVSEFSRMAIDPQRFVVIQNEBGLGPASPLDSTFYSLLEDDMDG 1014
 Db 957 KCMWIDSECRPREFELVSEFSRMAIDPQRFVVIQNEBGLGPASPLDSTFYSLLEDDMDG 1016
 QY 1015 LVDAEELVLPQQGFCFCDPAPAGAGWVHHRRSSSTSSGGGDLTLGLEBEEBAPSPLA 1074
 Db 1017 LVDAEELVLPQQGFCFCDPAPAGAGWVHHRRSSSTSSGGGDLTLGLEBEEBAPSPLA 1076
 QY 1075 PSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDVPLPSESTDYVAPLTCSPQ 1134
 Db 1077 PSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDVPLPSESTDYVAPLTCSPQ 1136
 QY 1135 PEYVNOPDVPRQPPSPREGPLPAARPAATLERAKTILSPKNGVYDVFAFGAVENPEY 1194
 Db 1137 PEYVNOPDVPRQPPSPREGPLPAARPAATLERAKTILSPKNGVYDVFAFGAVENPEY 1196
 QY 1195 LTPQGAAPOPHPPEPAPSPAFDNLXYWDOPPERGAPSPSTFKGPTTENPEYLGIDVPV 1253
 Db 1197 LTPQGAAPOPHPPEPAPSPAFDNLXYWDOPPERGAPSPSTFKGPTTENPEYLGIDVPV 1255
 RESULT 4
 AAU74545
 ID AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 XX 23-APR-2002 (first entry)
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KM Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KM anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KM stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KM thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KM glioma disorder; astrocytoma disorder; hypothalamic disorder;
 KM glandular disorder; macrophagal disorder; epithelial disorder;
 KM stromal disorder; blastococelic disorder; inflammatory disorder;
 KM angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 PN US2002001587-A1.
 PD 03-JAN-2002.
 PF 16-MAR-2001; 2001US-0811123.
 XX 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIM/) SLIKOWSKI M.
 XX
 PI Erickson S, Schwall R, Slikowski M;
 XX
 DR WPI; 2002-163686/21.
 DR N-PSDB; ABK14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 PT factor receptor; ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 PS Example 3; Fig 7; 93pp; English.
 XX
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytes, hypothalamic, glandular, macrophagal,
 CC epithelial, stromal, blastococelic, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 Query Match 97.0%; Score 6610; DB 23; Length 1255;
 Best Local Similarity 96.9%; Pred. No. 0;
 Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;
 QY 1 MELAALCRGILLALPPGAASVQCTGDMKRLPASPETHLDMLRHLVQCGVYQGNL 60
 Db 1 MELAALCRGILLALPPGAASVQCTGDMKRLPASPETHLDMLRHLVQCGVYQGNL 60
 QY 61 ELTYLPTNASLFLDIOIEVQGYVLIANNQVQVPLQRLIRIVRGTQLFEDYVALAVDNG 120
 Db 61 ELTYLPTNASLFLDIOIEVQGYVLIANNQVQVPLQRLIRIVRGTQLFEDYVALAVDNG 120
 QY 121 DPLNNTPTVTGASPGGLRELQRLSTELLKGVVLIOBNQQLCYQDTLLMKDIFHNKNOLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSTELLKGVVLIOBNQQLCYQDTLLMKDIFHNKNOLA 180
 QY 181 LTLIDTNRSRACHPCSPMCKGSRGWSESEDCSLRTVCAGGACRCKPLPTDCHEQC 240
 Db 181 LTLIDTNRSRACHPCSPMCKGSRGWSESEDCSLRTVCAGGACRCKPLPTDCHEQC 240
 QY 241 AAGCTGPKMSDCLACHFNHSGICELKCALVTYNTDFESHPNBEGRTPFASCTYACP 300
 Db 241 AAGCTGPKMSDCLACHFNHSGICELKCALVTYNTDFESHPNBEGRTPFASCTYACP 300
 QY 301 VYVLSIDVSGCTVCPLEHNOEYTAEDGTORCKSKPCARVCYGGMOYIKANSKFIGIT 360
 Db 301 VYVLSIDVSGCTVCPLEHNOEYTAEDGTORCKSKPCARVCYGGMOYIKANSKFIGIT 360
 QY 361 ELIEPAGCKKIFGSLAFLEPSFDGDPASNTAPLQPSQLQVFTLEBITGYLISAMPDSLP 420
 Db 361 IOEPAGCKKIFGSLAFLEPSFDGDPASNTAPLQPSQLQVFTLEBITGYLISAMPDSLP 420
 QY 421 DLSVFQNLQVIRGILHNGAVSLTQGLISWLGRLSRLRELGSGJALLIHNHNLGCVHTV 480
 Db 421 DLSVFQNLQVIRGILHNGAVSLTQGLISWLGRLSRLRELGSGJALLIHNHNLGCVHTV 480
 QY 481 PMDOLFNRNHALHTANRPEDECVGEGIALCHQLCARHGWGPGFTQCVNCSQFLRGQEC 540
 Db 481 PMDOLFNRNHALHTANRPEDECVGEGIALCHQLCARHGWGPGFTQCVNCSQFLRGQEC 540
 QY 541 VEECRVLOGLPREYVNAHCHLPCHCEQCPQNSVTCFGEADQCAVACHYKDPFCVANC 600
 Db 541 VEECRVLOGLPREYVNAHCHLPCHCEQCPQNSVTCFGEADQCAVACHYKDPFCVANC 600
 QY 601 PGGVPRDLSYMPIMKFPDEBAGCPCPINCHSCVDLDDKGPAGORASPLTSIYSAVVG 660
 Db 601 PGGVPRDLSYMPIMKFPDEBAGCPCPINCHSCVDLDDKGPAGORASPLTSIYSAVVG 660
 QY 661 ILLVVLGVVFGILL-----FNNFTVSFWLRVPPVSAASHLEPLTPSGAMPQAOAMRLK 714
 Db 661 ILLVVLGVVFGILL-----FNNFTVSFWLRVPPVSAASHLEPLTPSGAMPQAOAMRLK 714
 QY 715 ETELKRVKVLGAGAGTYKGIWIPDGENVKIPVAKIVLRENTSKAKKEIDEEYVWAG 774
 Db 715 ETELKRVKVLGAGAGTYKGIWIPDGENVKIPVAKIVLRENTSKAKKEIDEEYVWAG 774
 QY 717 ETELKRVKVLGAGAGTYKGIWIPDGENVKIPVAKIVLRENTSKAKKEIDEEYVWAG 776
 Db 717 ETELKRVKVLGAGAGTYKGIWIPDGENVKIPVAKIVLRENTSKAKKEIDEEYVWAG 776
 QY 775 VGGPYVSRLLGICLTSTVQVLTQMPYGCLLDHYRENRGRGSGODLNMWQIAKGSYL 834
 Db 775 VGGPYVSRLLGICLTSTVQVLTQMPYGCLLDHYRENRGRGSGODLNMWQIAKGSYL 834

```

DB 777 VGSFVSRLLIGICTSTVQLTQMLPGCLLDHVRERGRIGSDLLNMCQIAKMSYL 836
QY 835 EDVFLVHRDLAARVAVLVSPPHVKITDFGLARLDIDETEHADGKVPKIMMALESILR 894
DB 837 EDVFLVHRDLAARVAVLVSPPHVKITDFGLARLDIDETEHADGKVPKIMMALESILR 896
QY 895 RRTFGSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKEEFLPPPICTIDVYIMV 954
DB 897 RRTFGSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKEEFLPPPICTIDVYIMV 956
QY 955 KCMWIDSECRPFREIUSEFSRMAQDQRFVYIENEDLGPASPLDSTFFYSRLLEDDMGD 1014
DB 957 KCMWIDSECRPFREIUSEFSRMAQDQRFVYIENEDLGPASPLDSTFFYSRLLEDDMGD 1016
QY 1015 LVDAEYLVPOQGFCCDPAPAGAGMYHRRSSSTRSGGDLTLGLEPSEEAAPSPLA 1074
DB 1017 LVDAEYLVPOQGFCCDPAPAGAGMYHRRSSSTRSGGDLTLGLEPSEEAAPSPLA 1076
QY 1075 PSEGAGSTVFPGDILGMAKAGLQSLPTHDSPLORYSDDPTVPLPSETDGVVAFLTCSFQ 1134
DB 1077 PSEGAGSTVFPGDILGMAKAGLQSLPTHDSPLORYSDDPTVPLPSETDGVVAFLTCSFQ 1136
QY 1135 PEYVNPQPVRRPQPSPEPGPLPAPAPAGATLERAKTISPKNQGVKDVFAFGAVENPEY 1194
DB 1137 PEYVNPQPVRRPQPSPEPGPLPAPAPAGATLERAKTISPKNQGVKDVFAFGAVENPEY 1196
QY 1195 LTPQGGAAPQHPPEPAPFADNLVYWDQDPPEGAPSTFKGPTLENPEYLGIDVPV 1253
DB 1197 LTPQGGAAPQHPPEPAPFADNLVYWDQDPPEGAPSTFKGPTLENPEYLGIDVPV 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein, 1255 AA.
AC AAW01111;
XX
DT 01-JAN-1997 (first entry)
DE HER-2/neu protein.
XX
KM HER-2/neu; c-erbB1, p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
XX prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 676..1255
FT /label= intracellular_domain
FT /note= "claimed domain, useful for immunisation"
XX
PN MO9630514-A1.
XX
PD 03-OCT-1996.
XX
PF 28-MAR-1996; 96WO-US01689.
XX
PR 31-MAR-1995; 95US-041417.
XX
PA (UNITM) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI; 1996-455361/45.
XX
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
XX associated
XX
PS Claim 2; Page 56-61; 71pp; English.

```

```

XX CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6604; DB 17; Length 1255;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2.
QY 1 METALCRGGLLALLPFAASTQVCTGDMRLRPASPTHLDMLRHLVQCCQVQGNL 60
DB 1 MELALCRGGLLALLPFAASTQVCTGDMRLRPASPTHLDMLRHLVQCCQVQGNL 60
QY 61 ELTYLPTNASLFLQDICEVQGVLIANQVQVPLQRIYRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLFLQDICEVQGVLIANQVQVPLQRIYRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTSFSGSLRELOSLTEILKGVLIQBNPOLCYOPTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTSFSGSLRELOSLTEILKGVLIQBNPOLCYOPTILMKDIFHKNNOLA 180
QY 181 LTLIDTNRBACHPCSPMKSGRCWGESSEDCSLTRVYACAGCARCKPLPTDCHEQC 240
DB 181 LTLIDTNRBACHPCSPMKSGRCWGESSEDCSLTRVYACAGCARCKPLPTDCHEQC 240
QY 241 AAGCTGPKSDCLACHFNHSGICEHCPALVTYNTDFEEMNPBEGRTFPGASCTAEP 300
DB 241 AAGCTGPKSDCLACHFNHSGICEHCPALVTYNTDFEEMNPBEGRTFPGASCTAEP 300
QY 301 VNYLSTDVGSCTLVCELANQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKEITG 360
DB 301 VNYLSTDVGSCTLVCELANQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKEITG 360
QY 361 ELFFACCKKIFGSLALPESFGDDPAKSNAPLOPELOVFEETLEETGLYISAMPDISP 420
DB 361 IOFFACCKKIFGSLALPESFGDDPAKSNAPLOPELOVFEETLEETGLYISAMPDISP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSITLOGLGISWLGRLRELSGGLALIHNNHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSITLOGLGISWLGRLRELSGGLALIHNNHLCFVHTV 480
QY 481 PMDQFRNPHQALHTANRPEDECVGEGJACHQLCARGHCMWGPPTQCVNCSQFLRGQC 540
DB 481 PMDQFRNPHQALHTANRPEDECVGEGJACHQLCARGHCMWGPPTQCVNCSQFLRGQC 540
QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNSVTCFGEADQCVACAYKPPFCVANC 600
DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNSVTCFGEADQCVACAYKPPFCVANC 600
QY 601 PSGVKPDLSTYMPIMKPEDEGACQPCPINCTHSVDLDKGPAGGRASPITSISAVVG 660
DB 601 PSGVKPDLSTYMPIMKPEDEGACQPCPINCTHSVDLDKGPAGGRASPITSISAVVG 660
QY 661 ILLVVLGVVFGILL-----ENNFTVSFWLAVPKYSASHLEPLTSSGAMPQAOIRILK 714
DB 661 ILLVVLGVVFGILL-----ENNFTVSFWLAVPKYSASHLEPLTSSGAMPQAOIRILK 716
QY 715 ETELARKVYLGSAGFTVYKGIWIPDGENVKIIPVALKYRENTSPKANKELIDEAYVMAG 774
DB 717 ETELARKVYLGSAGFTVYKGIWIPDGENVKIIPVALKYRENTSPKANKELIDEAYVMAG 776
QY 775 VGSFVSRLLIGICTSTVQLTQMLPGCLLDHVRERGRIGSDLLNMCQIAKMSYL 834
DB 777 VGSFVSRLLIGICTSTVQLTQMLPGCLLDHVRERGRIGSDLLNMCQIAKMSYL 836

```

QY 835 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPKIMMALESILR 894
 DB 837 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPKIMMALESILR 896
 QY 895 RRFTHQSDVMSYCTTWELMTFGAKPYDGIARIEIPDLLEKGERLPPQPICTIDVYIMV 954
 DB 897 RRFTHQSDVMSYCTTWELMTFGAKPYDGIARIEIPDLLEKGERLPPQPICTIDVYIMV 956
 QY 955 KCMWIDSECRPRFRELVSERFARMADPQRFVIVIONEDLGPAAPLDSFTFYSILDDMDG 1014
 DB 957 KCMWIDSECRPRFRELVSERFARMADPQRFVIVIONEDLGPAAPLDSFTFYSILDDMDG 1016
 QY 1015 LVDAEEYLVPOQGFCCDPPAGAGAMVHHRSSSTSGGSDTLGLSEPEEAPRSPLA 1074
 DB 1017 LVDAEEYLVPOQGFCCDPPAGAGAMVHHRSSSTSGGSDTLGLSEPEEAPRSPLA 1076
 QY 1075 PSEGASDFVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSETDGVAPLTCSPQ 1134
 DB 1077 PSEGASDFVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSETDGVAPLTCSPQ 1136
 QY 1135 PEYVNPDPVRPQPSPEEGPLPAARPAATLERAKTLSPGKNGVYKDFAFGAVENPEY 1194
 DB 1137 PEYVNPDPVRPQPSPEEGPLPAARPAATLERAKTLSPGKNGVYKDFAFGAVENPEY 1196
 QY 1195 LTPGGGAAPQHPPEPAPFADNLVYWDQDPEERGAPESTFKGPTAENPEYGLDVPV 1253
 DB 1197 LTPGGGAAPQHPPEPAPFADNLVYWDQDPEERGAPESTFKGPTAENPEYGLDVPV 1255
 RESULT 6
 AAM92406
 ID AAM92406 standard; Protein: 1255 AA.
 AC AAM92406;
 XX 21-APR-1999 (first entry)
 DT 21-APR-1999 (first entry)
 XX Human HER-2/neu oncogene protein.
 DE Human HER-2/neu oncogene protein.
 XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 KW malignancy; treatment; tumour.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FT Region 676..1255
 FT /note="region which elicits immune response"
 XX
 XX US5869445-A.
 XX 09-FEB-1999.
 XX 01-APR-1996; 96US-0625101.
 XX 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 XX
 XX (UNIM) UNIV WASHINGTON.
 XX
 XX Cheever MA, Disis ML;
 XX WPI; 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX
 XX Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 XX
 XX Claim 3; Column 31-38; 26pp; English.
 XX
 XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or reoccurrence.
 CC
 XX

SQ Sequence 1255 AA;
 Query Match 96.9%; Score 6604; DB 20; Length 1255;
 Best Local Similarity 96.7%; Pred. No. 0;
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 MELALCRGGLLALPFGAASVQCTGDMKRLPASPEHLDMLRHLYOGQVQGNL 60
 DB 1 MELALCRGGLLALPFGAASVQCTGDMKRLPASPEHLDMLRHLYOGQVQGNL 60
 QY 61 ELTYLPTNLSLFLQDIQEVQGVYLIANQVQVPLQRLIRYRGTLFEDNYALVLDNG 120
 DB 61 ELTYLPTNLSLFLQDIQEVQGVYLIANQVQVPLQRLIRYRGTLFEDNYALVLDNG 120
 QY 121 DPLNNTPTVAGSPGGLRELQRLSLTEILKGGVLIQRPOLCYODITLWKDIFHKNOILA 180
 DB 121 DPLNNTPTVAGSPGGLRELQRLSLTEILKGGVLIQRPOLCYODITLWKDIFHKNOILA 180
 QY 181 LTLIDNRSRACHPCSPMKSGRCSGSESEDCQSLTRVYACGACRCKPLPTDCHEQC 240
 DB 181 LTLIDNRSRACHPCSPMKSGRCSGSESEDCQSLTRVYACGACRCKPLPTDCHEQC 240
 QY 241 AAGCTGPKKSDCLACLFHNSGICELHCAALVYNTDTEFSPNPEGRTPGASCTACP 300
 DB 241 AAGCTGPKKSDCLACLFHNSGICELHCAALVYNTDTEFSPNPEGRTPGASCTACP 300
 QY 301 YNVLSTDVSGCTLVCPBLHNOEYTAEDGTORCEKSKPCARVCYGLGMOYIKANSXFIGIT 360
 DB 301 YNVLSTDVSGCTLVCPBLHNOEYTAEDGTORCEKSKPCARVCYGLGMOYIKANSXFIGIT 360
 QY 361 ELIEFAGCKKIFGSLATLPSFDGDPASNTAPLOPELOVFEETLEETGLYISAMPDILP 420
 DB 361 IOEFAGCKKIFGSLATLPSFDGDPASNTAPLOPELOVFEETLEETGLYISAMPDILP 420
 QY 421 DLSVFONLQVIRGRILHNGAYSLTLOGLISWGLRLSELGSLALIHNNHLCVHRV 480
 DB 421 DLSVFONLQVIRGRILHNGAYSLTLOGLISWGLRLSELGSLALIHNNHLCVHRV 480
 QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARHCWGPPTQCVNCSQFLRGQEC 540
 DB 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARHCWGPPTQCVNCSQFLRGQEC 540
 QY 541 VECRPLQGLPREYVNAHRLCPCHPECOPONGSVTCFGEADOCVACAHYKDPFVCVAC 600
 DB 541 VECRPLQGLPREYVNAHRLCPCHPECOPONGSVTCFGEADOCVACAHYKDPFVCVAC 600
 QY 601 PSGVXPDLSTYMPIMKFPDEGACQPCPINCSTHSCVDLDDKGPASQASPLTISAVVG 660
 DB 601 PSGVXPDLSTYMPIMKFPDEGACQPCPINCSTHSCVDLDDKGPASQASPLTISAVVG 660
 QY 661 ILLVVLGVYFGILL-----FNNFTVSFWLAVPVKASASHEPLTPSGAMPQAOQMRILK 714
 DB 661 ILLVVLGVYFGILL-----FNNFTVSFWLAVPVKASASHEPLTPSGAMPQAOQMRILK 714
 QY 715 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKYLRENTSPKANKELIDEAYVMAG 774
 DB 715 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKYLRENTSPKANKELIDEAYVMAG 774
 QY 775 VGSFVYSRLGICLSTVQLVTQMLPYGGLDHTVRENTGRGLSGQDLNWCQAIAGMSYL 834
 DB 775 VGSFVYSRLGICLSTVQLVTQMLPYGGLDHTVRENTGRGLSGQDLNWCQAIAGMSYL 834
 QY 835 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPKIMMALESILR 894
 DB 837 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPKIMMALESILR 896

QY 895 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPTCTIDVYIMV 954
 DB 897 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPTCTIDVYIMV 956
 QY 955 KCMWIDSECRPRFRELIVSEFSRMARDPQRFVIVIONEDGSPASPLDSTYRSLEDDMDGD 1014
 DB 957 KCMWIDSECRPRFRELIVSEFSRMARDPQRFVIVIONEDGSPASPLDSTYRSLEDDMDGD 1016
 QY 1015 LVDAEEYLVPOQGFPCPDPAQAGMWHHRSSSTRSGGDLTLGLEPSEEAPRSPLA 1074
 DB 1017 LVDAEEYLVPOQGFPCPDPAQAGMWHHRSSSTRSGGDLTLGLEPSEEAPRSPLA 1076
 QY 1075 PSEGASDVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSENDGVAAPLTCSPO 1134
 DB 1077 PSEGASDVFDGDLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSENDGVAAPLTCSPO 1136
 QY 1135 PEYVNPDPVRPOPSPREGPLEAARPAAGATLERAKTISPGKNGVVKDVAFGAVENPEY 1194
 DB 1137 PEYVNPDPVRPOPSPREGPLEAARPAAGATLERAKTISPGKNGVVKDVAFGAVENPEY 1196
 QY 1195 LTPGGGAAPQHPHPPAFSPAFDNLVYWDODPPEEGAPSTFKGTPTAENPEYLG.DVAV 1253
 DB 1197 LTPGGGAAPQHPHPPAFSPAFDNLVYWDODPPEEGAPSTFKGTPTAENPEYLG.DVAV 1255

RESULT 7

1D AAB21198 standard; protein; 1255 AA.

XX AAB21198;

DT 12-JAN-2001 (first entry)

DE Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;

KM breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX colon cancer.

XX Homo sapiens.

XX MO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

XX (SMK) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX WPI: 2000-505976/45.

XX N-PSDB; AAB89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins

XX useful for vaccinating against breast, ovarian, colon, lung and

XX prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.

XX Sequence 1255 AA;

SQ Query Match 96.9%; Score 6604; DB 21; Length 1255;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 MELALCRNGLIALLPAGASTQVCTGDMKRLPASETHLDMRLHYOGQVQVGNL 60
 DB 1 MELALCRNGLIALLPAGASTQVCTGDMKRLPASETHLDMRLHYOGQVQVGNL 60
 QY 61 ELTYLPTNLSLSTLOIOEVQGVLLAHNOVQVPLRQRLRYRGTLDFEDNVALAVLNG 120
 DB 61 ELTYLPTNLSLSTLOIOEVQGVLLAHNOVQVPLRQRLRYRGTLDFEDNVALAVLNG 120
 QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYQDTILMKDIFHKNNOLA 180
 DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRNPOLCYQDTILMKDIFHKNNOLA 180
 QY 181 LTLIDNRSRACHPCSPMKGSRCWESSEDCOSLRTVCAAGCACKGPLPTDCHBQC 240
 DB 181 LTLIDNRSRACHPCSPMKGSRCWESSEDCOSLRTVCAAGCACKGPLPTDCHBQC 240
 QY 241 AACCTGPKASDCLACLHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGASCYTACP 300
 DB 241 AACCTGPKASDCLACLHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGASCYTACP 300
 QY 301 YNTLSTDVSGCTLVCELNQEVTAEBGTQRCCKSPCARVCYGLCMQYIKANSKFTGIT 360
 DB 301 YNTLSTDVSGCTLVCELNQEVTAEBGTQRCCKSPCARVCYGLCMQYIKANSKFTGIT 360
 QY 361 ELFEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEOLQVFEETLEITGYLYISAMPDSL 420
 DB 361 ELFEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEOLQVFEETLEITGYLYISAMPDSL 420
 QY 421 DLSVFNOLQVIRRIIHNGAYSLTLOGLSISWGLSLRLSGGLALIHNTLCEVHTV 480
 DB 421 DLSVFNOLQVIRRIIHNGAYSLTLOGLSISWGLSLRLSGGLALIHNTLCEVHTV 480
 QY 481 PMDOLFRNPQALHTANRPEDCEVSEGLACHOLCARGHCWGPGPQVNCOSFLRGQEC 540
 DB 481 PMDOLFRNPQALHTANRPEDCEVSEGLACHOLCARGHCWGPGPQVNCOSFLRGQEC 540
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 QY 601 PSGVKPDLSTYMPIMKPFDEEGACQPCGICNTHSCVDLDDKCPAEQASPLTISIIVVG 660
 DB 601 PSGVKPDLSTYMPIMKPFDEEGACQPCGICNTHSCVDLDDKCPAEQASPLTISIIVVG 660
 QY 661 ILVVVLGVVFFGLI-----FNNFTVSFVLRYKVSASHLBEPLTSGAMPQNAQRILK 714
 DB 661 ILVVVLGVVFFGLI-----FNNFTVSFVLRYKVSASHLBEPLTSGAMPQNAQRILK 714
 QY 715 ETELARKKYLGSAGFGVYKGIWIPGGENYKIVALKVLRBNTSPANKKILDEAYVMAG 774
 DB 715 ETELARKKYLGSAGFGVYKGIWIPGGENYKIVALKVLRBNTSPANKKILDEAYVMAG 774
 QY 717 ETELARKKYLGSAGFGVYKGIWIPGGENYKIVALKVLRBNTSPANKKILDEAYVMAG 776
 DB 717 ETELARKKYLGSAGFGVYKGIWIPGGENYKIVALKVLRBNTSPANKKILDEAYVMAG 776
 QY 775 VGSPPYVSRLLGICLTSTVQLVLTQIMPEYGLLDHVRNRRGLSGODLLNMCMQIAKMSYL 834
 DB 775 VGSPPYVSRLLGICLTSTVQLVLTQIMPEYGLLDHVRNRRGLSGODLLNMCMQIAKMSYL 834
 QY 835 EDVRLVHDLAANNVLKSPNHVKITDFGLARLLIDETBYHADGKVPKIMVALSILR 894
 DB 835 EDVRLVHDLAANNVLKSPNHVKITDFGLARLLIDETBYHADGKVPKIMVALSILR 894
 QY 895 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPTCTIDVYIMV 954
 DB 897 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPQPTCTIDVYIMV 956
 QY 955 KCMWIDSECRPRFRELIVSEFSRMARDPQRFVIVIONEDGSPASPLDSTYRSLEDDMDGD 1014


```

Db      957 KCMWIDSECRPRFELVSEFSRMARDPQRFVVIQNEDELGSAPLDSFTFYSLLEDDMGD 1016
Qy      1015 LVDAEYLVPQOGFFCDDPAPGAGMVAHRRSSSTRSGGDLTLGLEBEEBAPRSPILA 1074
Db      1017 LVDAEYLVPQOGFFCDDPAPGAGMVAHRRSSSTRSGGDLTLGLEBEEBAPRSPILA 1076
Qy      1075 PSBGAGSDVDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSESTDGYVAPLTCSPQ 1134
Db      1077 PSBGAGSDVDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSESTDGYVAPLTCSPQ 1136
Qy      1135 PEYVQDPVRRPQPSPEEGPLPAARPAAGATLERPKTSLPGKNGVYKDVAFAGAVENPEY 1194
Db      1137 PEYVQDPVRRPQPSPEEGPLPAARPAAGATLERPKTSLPGKNGVYKDVAFAGAVENPEY 1196
Qy      1195 LTPGGAPQPHPPAPSPAFDNLVYWDQDPPEEGAPSPFTFKGPTAENPEYLGLDVVP 1253
Db      1197 LTPGGAPQPHPPAPSPAFDNLVYWDQDPPEEGAPSPFTFKGPTAENPEYLGLDVVP 1255

```

RESULT 8
AAy84780
ID AAY84780 standard; Protein; 1255 AA.

AC AAY84780;
DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the SPICE erbB-2 receptor protein.

KM SPICE erbB-2 receptor protein; cell transformation disorder; cancer;
KM tumor cell proliferation; tissue degeneration; arthropathy;
KM bone resorption; inflammatory disease; degenerative disorder;
KM wound healing.

OS Homo sapiens.

PN MO200020579-A1.

PD 13-APR-2000.

PF 01-OCT-1999; 99MO-CA00912.

PR 02-OCT-1998; 98US-0165192.

PA (UYMC-) UNIV MCMASTER.

PI Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

DR N-PSDB; AAA14812.

PT Nucleic acid encoding an erbB 2 receptor protein designated SPICE

PS Claim 3; Fig 2; 60pp; English.

CC The present sequence represents a SPICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.

Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 21; Length 1255;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

```

Qy      1 MELAALCRWGLLJALLPPGAASYCTGTDMKRLPASBETHLMLRHLVQCGVYVQNTL 60
Db      1 MELAALCRWGLLJALLPPGAASYCTGTDMKRLPASBETHLMLRHLVQCGVYVQNTL 60
Qy      61 ELTYLPTNASSFLQDIOEVQGVYLAHQVQVQLQRLIRVGRQLFEDNALAVLDNG 120
Db      61 ELTYLPTNASSFLQDIOEVQGVYLAHQVQVQLQRLIRVGRQLFEDNALAVLDNG 120
Qy      121 DPLNNTPTVTCASPGSLBELQRLSTELIKGVLLIQRPOLCYQDTILMKDIFHNQOLA 180
Db      121 DPLNNTPTVTCASPGSLBELQRLSTELIKGVLLIQRPOLCYQDTILMKDIFHNQOLA 180
Qy      181 LTLIDTNRARACHPCSPKCKSGRCKGSESDCOSLTRIVCAGGACARCKPLPTDCHEQC 240
Db      181 LTLIDTNRARACHPCSPKCKSGRCKGSESDCOSLTRIVCAGGACARCKPLPTDCHEQC 240
Qy      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRTFGASCVTACP 300
Db      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRTFGASCVTACP 300
Qy      301 YNYLSTDVGSCTLVCPLNQVETADGTQRCCKSPCARVCYGLGMQYIKANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPLNQVETADGTQRCCKSPCARVCYGLGMHLEBRAVAVTSAN 360
Qy      361 ELIEFAGCKKIFESLAFLESFPGDPASTAPLQEPQLQVFELEITGLVYISAMPDSIP 420
Db      361 IOEFAGCKKIFESLAFLESFPGDPASTAPLQEPQLQVFELEITGLVYISAMPDSIP 420
Qy      421 DLSVFQNLQVIRGILHNGAVSLTLQIGISLGLRSRLRELSGLALIHNNHLFCVHTV 480
Db      421 DLSVFQNLQVIRGILHNGAVSLTLQIGISLGLRSRLRELSGLALIHNNHLFCVHTV 480
Qy      481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCMGPGFTQCVNCSQIRGQBC 540
Db      481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCMGPGFTQCVNCSQIRGQBC 540
Qy      541 VEECRVLQGLPREYVNAHRCPLCHPECCQPNOSVTCFGEADQCAACHYKDPFCVABC 600
Db      541 VEECRVLQGLPREYVNAHRCPLCHPECCQPNOSVTCFGEADQCAACHYKDPFCVABC 600
Qy      601 PSQVNPDSLYMPIWKFPEDEGACQPCPINCTHSCVDLDDKCPAEQORASPLTISAVVG 660
Db      601 PSQVNPDSLYMPIWKFPEDEGACQPCPINCTHSCVDLDDKCPAEQORASPLTISAVVG 660
Qy      661 ILVVVLGVVFGILL-----FNNFTVSFMLRVPKVSASHLEPLTPSGAMPQAOQMRILK 714
Db      661 ILVVVLGVVFGILLKRRQOKIRKYMRLLEQETEL---VEPLTPSGAMPQAOQMRILK 716
Qy      715 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKEIDEAYVYAG 774
Db      715 ETELKRVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKYLRNTSPKANKEIDEAYVYAG 776
Qy      775 VCSPIVSRILGICLSTVQVLTQMLPYGCLLDHVENRGLSGQDLVWCMQIAKMSYL 834
Db      775 VCSPIVSRILGICLSTVQVLTQMLPYGCLLDHVENRGLSGQDLVWCMQIAKMSYL 836
Qy      835 EGVRLVHMDLAARNVLYVSPNHVKITDGLARLIDIDEVEYADGKPIKMALESLIR 894
Db      835 EGVRLVHMDLAARNVLYVSPNHVKITDGLARLIDIDEVEYADGKPIKMALESLIR 896
Qy      895 RRFTHQSDVMSGYVTWELMTFGAXYDGIPIAREIPDLLERGERLPQPICTIDVYIMV 954
Db      895 RRFTHQSDVMSGYVTWELMTFGAXYDGIPIAREIPDLLERGERLPQPICTIDVYIMV 956
Qy      955 KCMWIDSECRPRFELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYSLLEDDMGD 1014
Db      955 KCMWIDSECRPRFELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYSLLEDDMGD 1016
Qy      1015 LVDAEYLVPQOGFFCDDPAPGAGMVAHRRSSSTRSGGDLTLGLEBEEBAPRSPILA 1074

```

DB 1017 LVDAEYLVPQGFCCDPAPAGAGWVHHRSSSTSGGGDLTLGLEPSEEEAPRSPPLA 1076
 QY 1075 PSEAGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSPSETDGVAPLTCSPQ 1134
 DB 1077 PSEAGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSPSETDGVAPLTCSPQ 1136
 QY 1135 PEYVNOPDVPRPQPSPREGGLPAARPAAGATLERAKTILSPGKNGYVVKVFAFGAVENPEY 1194
 DB 1137 PEYVNOPDVPRPQPSPREGGLPAARPAAGATLERAKTILSPGKNGYVVKVFAFGAVENPEY 1196
 QY 1195 LTPQGAAPQHPHPAPAPADNLYYWDODPPERGAPSPFPGTPTAENPEYLGLDVPPV 1253
 DB 1197 LTPQGAAPQHPHPAPAPADNLYYWDODPPERGAPSPFPGTPTAENPEYLGLDVPPV 1255

RESULT 9

AAB85458 ID AAB85458 standard; Protein; 1255 AA.

AC AAB85458;

DT 25-SEP-2001 (first entry)

DE Human HER-2/neu protein.

KM Antigen-presenting cell; immunogenic; immune response; HER-2/neu; oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.

OS Homo sapiens.

PN WO200153463-A2.

PD 26-JUL-2001.

PF 19-JAN-2001; 2001WO-US01850.

PR 21-JAN-2000; 2000US-0177545.

PA (CORI-) CORIXA CORP.

PI Cheever MA, Hand-Zimmermann S;

DR WPI; 2001-476112/51.

DR N-PSDB; AAH23392.

PT New antigen-presenting cells, useful as vaccines for eliciting or enhancing an immune response to HER-2/neu protein, particularly useful for treating or preventing cancer, e.g. breast cancer

PS Claim 2; Page 41-46; 49pp; English.

XX The invention provides an isolated antigen-presenting cell, which expresses at least an immunogenic portion of a polypeptide that produces an immune response to HER-2/neu protein. The antigen-presenting cells are useful as vaccines for eliciting or enhancing an immune response to CC HER-2/neu protein, particularly in treating or preventing malignancies in which the HER-2/neu oncogene is associated. Specifically, these are useful for treating or preventing cancer, e.g. breast cancer, ovarian, colon, lung or prostate cancers. The present sequence represents the human HER-2/neu protein (also known as p185 or c-erbB2).

XX Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 22; Length 1255;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1318; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 MELALACRMGLLALLPFGAASSTVCTGTDMLRLPASPETHLDMRLHYOGCCVYVGNL 60
 DB 1 MELALACRMGLLALLPFGAASSTVCTGTDMLRLPASPETHLDMRLHYOGCCVYVGNL 60
 QY 61 ELTYLPTNASTSLFLODIOEVQGVLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

DB 61 ELTYLPTNASTSLFLODIOEVQGVLIHNVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVGTASPGGLRELQRLSTLEILKGVLIQRNEPOLCYDTIIMKDI FHKNNQLA 180
 DB 121 DPLNNTPTVGTASPGGLRELQRLSTLEILKGVLIQRNEPOLCYDTIIMKDI FHKNNQLA 180
 QY 181 LTLIDNRSRACHPGSPMKSGRCMGESSEDDQSLTRVYAGAGCARCKRPLPTDCHEQC 240
 DB 181 LTLIDNRSRACHPGSPMKSGRCMGESSEDDQSLTRVYAGAGCARCKRPLPTDCHEQC 240
 QY 241 AAGCTPKASDCLACLFHNSGICELHCPALVTYNTDTPESMPNEGRYTFGASCTYACP 300
 DB 241 AAGCTPKASDCLACLFHNSGICELHCPALVTYNTDTPESMPNEGRYTFGASCTYACP 300
 QY 301 YNYLSTDVSGCTLVCPFLNQBYTADBDTQRCCKSPKARVCYGLGMQYIKANSKRTIGIT 360
 DB 301 YNYLSTDVSGCTLVCPFLNQBYTADBDTQRCCKSPKARVCYGLGMQYIKANSKRTIGIT 360
 QY 361 ELFFAGCKKIFGSLAFPLPESFGDDPASNTAPLOPELOQVFEETLEETGYLYISAMPDSTP 420
 DB 361 ELFFAGCKKIFGSLAFPLPESFGDDPASNTAPLOPELOQVFEETLEETGYLYISAMPDSTP 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSLRELSSGLALIHNNTHLCFVHTV 480
 DB 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSLRELSSGLALIHNNTHLCFVHTV 480
 QY 481 PMQULFRNPHQALHNTANPEDECEVEGLACHOLCARGHCMKPGPQVNCGOLFAGQGC 540
 DB 481 PMQULFRNPHQALHNTANPEDECEVEGLACHOLCARGHCMKPGPQVNCGOLFAGQGC 540
 QY 541 VEECRVLQGLPREYVNAARHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKDPFCVAC 600
 DB 541 VEECRVLQGLPREYVNAARHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKDPFCVAC 600
 QY 601 PSGVXPDLSTYMPWKPPDEEGACQCPINCTHSQVLDLDRKGAERASPLSTVAVVG 660
 DB 601 PSGVXPDLSTYMPWKPPDEEGACQCPINCTHSQVLDLDRKGAERASPLSTVAVVG 660
 QY 661 ILVVLVGLGVFGILIKRQOKIRKYMRLQCTEL----VEPLTSGAMPNQAQRIILK 714
 DB 661 ILVVLVGLGVFGILIKRQOKIRKYMRLQCTEL----VEPLTSGAMPNQAQRIILK 714
 QY 715 ETELARKVYLGSAFSTVYKGIWIPDGENVKIPVAIKVRENTSPRANKIIDEAYVAG 774
 DB 715 ETELARKVYLGSAFSTVYKGIWIPDGENVKIPVAIKVRENTSPRANKIIDEAYVAG 774
 QY 775 VGSPPVSRLLGLTSTVQLVQLMPEYGLLDHVRNBRGLSODLLNCOMQIAKMSYL 834
 DB 775 VGSPPVSRLLGLTSTVQLVQLMPEYGLLDHVRNBRGLSODLLNCOMQIAKMSYL 834
 QY 835 EDVRLVHRDLAARNVLVKSBNHYKINDFGARLLDIDETRYADGGKVPIMKMALESILR 894
 DB 835 EDVRLVHRDLAARNVLVKSBNHYKINDFGARLLDIDETRYADGGKVPIMKMALESILR 894
 QY 895 RRTTHOSDWVSQVYTWELMTFGAKYDGIIPAEIIDLLEKGRLLQOPPCITDVMYMW 954
 DB 895 RRTTHOSDWVSQVYTWELMTFGAKYDGIIPAEIIDLLEKGRLLQOPPCITDVMYMW 954
 QY 955 KCMWISSECPREELVSEFSRMAPRQPFVVIQNEDLGPASPLDSTFYRSLLEDDMDG 1014
 DB 955 KCMWISSECPREELVSEFSRMAPRQPFVVIQNEDLGPASPLDSTFYRSLLEDDMDG 1014
 QY 1015 LVDAEYLVPQGFCCDPAPAGAGWVHHRSSSTSGGGDLTLGLEPSEEEAPRSPPLA 1074
 DB 1015 LVDAEYLVPQGFCCDPAPAGAGWVHHRSSSTSGGGDLTLGLEPSEEEAPRSPPLA 1074
 QY 1075 PSEAGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSPSETDGVAPLTCSPQ 1134
 DB 1075 PSEAGSDVFDGDLGMAAKGLOSLPTHDPSPLOQRYSEDPVLPSPSETDGVAPLTCSPQ 1136
 QY 1135 PEYVNOPDVPRPQPSPREGGLPAARPAAGATLERAKTILSPGKNGYVVKVFAFGAVENPEY 1194

Db 1137 PEYVNPQPVRRQGPSREGGPLPAARPAACATTLERPTLSPGKGVVXVDYFAFGAVENBEY 1196

QY 1195 LTPOGGAAPQHPPEAFSPAFNDLYWMDQPPREGCAPSTFGKPTTAENPEYLGLDVPV 1253

Db 1197 LTPOGGAAPQGHPRPAFSPAFNDLYWMDQPPREGCAPSTFGKPTTAENPEYLGLDVPV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA

AC AAG88267;

DT 11-SEP-2001 (first entry)

HER2/neu amino acid sequence.

KM Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KM immune response; vaccine; tumour; cancer; cytosolic; immunostimulant
 KM tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
 XX
 OS Homo sapiens.

PN WO200141787-A1

PD 14-JUN-2001

PF 11-DEC-2000; 2000WO-US33591

PR 10-DEC-1999; 99US-0458299

PA (EPIM-) EPIMUNE INC

PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;

DR WPI; 2001-374995/39

PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer
XX
ES Disclosure, Page 15, 1999p, English.

The present invention describes isolated prepared HER2/neu epitopes (1). Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is culture in vitro and binds to a complex of an epitope (I), bound to a human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I) and a second epitope and the peptide is less than 50 contiguous amino acids that have 100% identity with a native peptide sequence of HER2/neu; (3) a vaccine composition (III) comprising (II) and a pharmaceutical excipient; (4) an isolated nucleic acid encoding a peptide comprising (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic and immunostimulant activities, and can be used in vaccines. (I), (II) and (III) are useful for inducing cellular immune responses for the prevention and treatment of cancer. (I) and (II) are useful for monitoring or evaluating an immune response to a tumour-associated antigen when incubated with a T lymphocyte sample from a patient and detecting the presence of bound T lymphocyte to (I) or (II). Epitope based vaccines mean that immunosuppressive epitopes that may be present in whole antigens may be avoided. Selected epitopes may be combined to enhance immunogenicity. The possible pathological side effects caused by infectious agents or whole protein antigen is eliminated. The vaccine provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Epitope-based anti-tumour vaccines provides the opportunity to combine epitopes derived from multiple tumour-associated molecules addressing the problem of tumour-tumour variability and reducing the likelihood of tumour escape due to antigen loss. AAG8266 to AAG8921 represent amino acid sequences used in the exemplification of the present invention.

SQ Sequence 1255 AA;

Query Match	96.98; Score 6604; DB 22; Length 1255;
-------------	--

Best Local Similarity 96.7%; Pred. No. 0;
Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2

QY	1	MELAA	CRMG	LLALL	PPGA	STOV	CTDM	KL	LP	SP	PTH	DM	RL	YOG	CCV	YV	NL	60																																						
Db	1	MELAA	CRMG	LLALL	PPGA	STOV	CTDM	KL	LP	SP	PTH	DM	RL	YOG	CCV	YV	NL	60																																						
QY	61	ELTYL	PNTAS	LS	F	ODI	OE	VGYV	L	AH	NO	YV	P	LO	R	L	IV	RG	TO	LE	ED	NYA	L	VD	NG	120																														
Db	61	ELTYL	PNTAS	LS	F	ODI	OE	VGYV	L	AH	NO	YV	P	LO	R	L	IV	RG	TO	LE	ED	NYA	L	VD	NG	120																														
QY	121	DLNNT	TPV	GA	SG	RE	LO	BS	TE	IL	K	SG	V	L	IO	NP	OU	C	YD	OT	IM	K	O	I	PH	NO	LA	180																												
Db	121	DLNNT	TPV	GA	SG	RE	LO	BS	TE	IL	K	SG	V	L	IO	NP	OU	C	YD	OT	IM	K	O	I	PH	NO	LA	180																												
QY	181	LTLID	TR	SR	ACH	PC	S	PM	CK	S	R	C	M	G	S	S	E	D	C	O	S	L	TR	V	C	A	G	CA	R	K	GP	L	PT	D	C	HE	OC	240																		
Db	181	LTLID	TR	SR	ACH	PC	S	PM	CK	S	R	C	M	G	S	S	E	D	C	O	S	L	TR	V	C	A	G	CA	R	K	GP	L	PT	D	C	HE	OC	240																		
QY	241	AAGCT	GR	H	S	C	L	A	C	L	H	P	N	S	G	I	E	L	H	C	P	A	L	V	T	N	T	D	T	E	S	E	M	P	E	G	Y	T	T	G	A	S	C	V	T	A	C	P	300							
Db	241	AAGCT	GR	H	S	C	L	A	C	L	H	P	N	S	G	I	E	L	H	C	P	A	L	V	T	N	T	D	T	E	S	E	M	P	E	G	Y	T	T	G	A	S	C	V	T	A	C	P	300							
QY	301	YNY	ST	D	V	G	S	T	L	C	P	L	H	N	O	E	V	T	A	E	D	T	O	R	C	E	K	S	K	P	C	A	R	C	A	R	C	Y	G	M	O	Y	I	A	N	S	F	I	G	T	360					
Db	301	YNY	ST	D	V	G	S	T	L	C	P	L	H	N	O	E	V	T	A	E	D	T	O	R	C	E	K	S	K	P	C	A	R	C	A	R	C	Y	G	M	O	Y	I	A	N	S	F	I	G	T	360					
QY	361	EL	E	P	A	G	C	K	I	F	G	S	I	A	P	E	S	F	D	G	P	A	S	T	A	P	L	O	E	O	L	O	V	F	E	T	T	E	I	T	G	L	Y	L	I	S	A	M	P	D	S	L	P	420		
Db	361	EL	E	P	A	G	C	K	I	F	G	S	I	A	P	E	S	F	D	G	P	A	S	T	A	P	L	O	E	O	L	O	V	F	E	T	T	E	I	T	G	L	Y	L	I	S	A	M	P	D	S	L	P	420		
QY	421	DL	S	F	O	N	L	O	V	R	G	I	L	N	G	A	S	L	T	O	G	I	S	W	L	G	R	S	T	R	E	L	G	S	G	A	L	I	H	N	T	H	L	C	F	N	T	V	480							
Db	421	DL	S	F	O	N	L	O	V	R	G	I	L	N	G	A	S	L	T	O	G	I	S	W	L	G	R	S	T	R	E	L	G	S	G	A	L	I	H	N	T	H	L	C	F	N	T	V	480							
QY	481	F	M	D	L	F	R	N	P	O	A	L	L	H	T	A	N	R	P	E	D	E	C	V	G	E	L	A	C	H	O	L	C	A	R	G	H	C	M	G	P	G	T	O	C	V	N	S	O	F	L	R	G	E	C	540
Db	481	F	M	D	L	F	R	N	P	O	A	L	L	H	T	A	N	R	P	E	D	E	C	V	G	E	L	A	C	H	O	L	C	A	R	G	H	C	M	G																

Db 1017 LVDAEEYLVPOQGFPCDPAAGAGMHHHRSSSTSGGGDLTLGLEPSEEEAPRSPLA 1076
 Qy 1075 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDEPTVPLPSETDGVVAPLTCSPQ 1134
 Db 1077 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDEPTVPLPSETDGVVAPLTCSPQ 1136
 Qy 1135 PEYVNOPDVRPQPPSPREGPLPAARPAATLIERAKTILSPKNGVYKXVAFAGAVENPEY 1194
 Db 1137 PEYVNOPDVRPQPPSPREGPLPAARPAATLIERAKTILSPKNGVYKXVAFAGAVENPEY 1196
 Qy 1195 LTPQGAAPQHPHPAFSPAFDNLVYWDQDPPEGAPSPSTFKGTPTAENPEYLGLDVYV 1253
 Db 1197 LTPQGAAPQHPHPAFSPAFDNLVYWDQDPPEGAPSPSTFKGTPTAENPEYLGLDVYV 1255

RESULT 11

AAE24067
 ID AAE24067 standard; Protein; 1255 AA.

AAE24067;

23-SEP-2002 (first entry)

Human Her-2 protein.

Human; Her-2; epidermal growth factor receptor 2; infection; cancer;

hyperproliferative disorder; prophylaxis; inflammation; antisense;

tumour; gene therapy; phosphothioate backbone.

Human sapiens.

WO200222636-A1.

21-MAR-2002.

12-SEP-2001; 2001WO-US28572.

15-SEP-2000; 2000US-0663834.

(ISIS-) ISIS PHARM INC.

Bennett CF, Cowase LM;

WPI; 2002-471192/50.

N-PSDB; AAD38904.

Novel antisense oligonucleotide which modulates the expression of Human

Epidermal Growth Factor receptor, Her2, is useful for treating tumors

inflammation or to prevent infection in humans -

Example 13; Page 95-107; 116pp; English.

The invention relates to antisense compounds targeted to a nucleic

acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)

that specifically hybridises with and inhibits the expression of Her2.

Antisense compounds of the invention are used for treating diseases or

conditions associated with Her2 such as hyperproliferative disorders

e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,

neural or cardiac cancer. They are also useful prophylactically e.g.

to prevent or delay infection, inflammation and tumour formation. The

invention is also used in gene therapy. The present sequence is human

Her-2 protein.

Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 23; Length 1255;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

Qy 1 MELAAICRWGLLALILPQGAATGYCTGDMKRLPASPETHLDWLRLHYGGCCVYQGNL 60

Db 1 MELAAICRWGLLALILPQGAATGYCTGDMKRLPASPETHLDWLRLHYGGCCVYQGNL 60

Qy 61 ELTYLEPTNASTSLTODICEVQGVYLIAHNOVQVPLQRLRIYRGTOLEEDNYALAVLDNG 120
 Db 61 ELTYLEPTNASTSLTODICEVQGVYLIAHNOVQVPLQRLRIYRGTOLEEDNYALAVLDNG 120
 Qy 121 DPLANTTPTVYAGSPGGLARELOLSLEILKGVLLIQRNPOLCYODITLWKDI FHKNNOLA 180
 Db 121 DPLANTTPTVYAGSPGGLARELOLSLEILKGVLLIQRNPOLCYODITLWKDI FHKNNOLA 180
 Qy 181 LTLIDNRSBACHPGSPMKCKSGCWESSEDCOSLTRTYCAGGACAKGRLPTDCHECQ 240
 Db 181 LTLIDNRSBACHPGSPMKCKSGCWESSEDCOSLTRTYCAGGACAKGRLPTDCHECQ 240
 Qy 241 AACCTGPKHSDDCLACHFNHSGICEJHCALVTYNTDTESMNPFGRTYFGASCYAC 300
 Db 241 AACCTGPKHSDDCLACHFNHSGICEJHCALVTYNTDTESMNPFGRTYFGASCYAC 300
 Qy 301 YNYLSTDVSGCTLVCEPLHNOEYTAEDGTQCEKSPCARVCYGLMOQYIKANSKITGIT 360
 Db 301 YNYLSTDVSGCTLVCEPLHNOEYTAEDGTQCEKSPCARVCYGLMOQYIKANSKITGIT 360
 Qy 361 ELFEAGCKXIFGSLAFLESPFODGPASNTAPLOPELOVETLEETITGYLYISAMPDSL 420
 Db 361 ELFEAGCKXIFGSLAFLESPFODGPASNTAPLOPELOVETLEETITGYLYISAMPDSL 420
 Qy 421 DLSEFQNLQYTRGRILHNGAYSLTLQGLISWLGRLSLEBLSGLALIHNTLCEVHTV 480
 Db 421 DLSEFQNLQYTRGRILHNGAYSLTLQGLISWLGRLSLEBLSGLALIHNTLCEVHTV 480
 Qy 481 PMQULFRNPQALLHFNANPEDECVESGLACHOLCARGHCMGPPQCNVCSQFLGQEC 540
 Db 481 PMQULFRNPQALLHFNANPEDECVESGLACHOLCARGHCMGPPQCNVCSQFLGQEC 540
 Qy 541 VECRVLQGLPREYVNAHRLCPHPCCOPONGSVTCFGEADCCVACAHKDPFCVARC 600
 Db 541 VECRVLQGLPREYVNAHRLCPHPCCOPONGSVTCFGEADCCVACAHKDPFCVARC 600
 Qy 601 PSGVKPDLSTWPIWKPPDEEGACQPCPINCSTCVLDLDDKCAEORASFLTSVAVVG 660
 Db 601 PSGVKPDLSTWPIWKPPDEEGACQPCPINCSTCVLDLDDKCAEORASFLTSVAVVG 660
 Qy 661 ILVYVGVVFGILIKRQOKIRKYMRLDQTEL---VEPLTSGAMPNQAQRILK 714
 Db 661 ILVYVGVVFGILIKRQOKIRKYMRLDQTEL---VEPLTSGAMPNQAQRILK 714
 Qy 715 ETELKRVKVLGSGAFGVYKGIWIPGENYKIPVAIKVRENTSPANKELDEAYVMAG 774
 Db 715 ETELKRVKVLGSGAFGVYKGIWIPGENYKIPVAIKVRENTSPANKELDEAYVMAG 774
 Qy 775 VGSPPYVRLIGICLTSTVQLVQLOLMEYGCILDHVRNRRGLSGODLLNMCMQIAKMSYL 834
 Db 775 VGSPPYVRLIGICLTSTVQLVQLOLMEYGCILDHVRNRRGLSGODLLNMCMQIAKMSYL 834
 Qy 835 EDRLVARDLAANVLVKSFNHVKITDFGLARLJDIDETEHADGKVPKXMALESILR 894
 Db 835 EDRLVARDLAANVLVKSFNHVKITDFGLARLJDIDETEHADGKVPKXMALESILR 894
 Qy 895 RRTTHOSDVSYGTTWELMTFGAKXYDGIIPAEIPDLLEKGRLLQPPCTIDVYMIWY 954
 Db 895 RRTTHOSDVSYGTTWELMTFGAKXYDGIIPAEIPDLLEKGRLLQPPCTIDVYMIWY 954
 Qy 955 KCMNIDECRPRRELVSERFARMAPQRFVYQNEDLGPASLDSTFYSLIEDDMGD 1014
 Db 955 KCMNIDECRPRRELVSERFARMAPQRFVYQNEDLGPASLDSTFYSLIEDDMGD 1014
 Qy 1015 LVDAEEYLVPOQGFPCDPAAGAGMHHHRSSSTSGGGDLTLGLEPSEEEAPRSPLA 1074
 Db 1015 LVDAEEYLVPOQGFPCDPAAGAGMHHHRSSSTSGGGDLTLGLEPSEEEAPRSPLA 1074
 Qy 1075 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDEPTVPLPSETDGVVAPLTCSPQ 1134
 Db 1075 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDEPTVPLPSETDGVVAPLTCSPQ 1134
 Qy 1135 PEYVNOPDVRPQPPSPREGPLPAARPAATLIERAKTILSPKNGVYKXVAFAGAVENPEY 1194
 Db 1135 PEYVNOPDVRPQPPSPREGPLPAARPAATLIERAKTILSPKNGVYKXVAFAGAVENPEY 1194

Db 1137 PEYVNGPDSVSPQPSREBGLPAPAPGATLERPKTSLPGNGVVKVQVPAFGAVENPEY 1196
 QY 1195 LTFPGGAAPQHPHPAPSPAFNDLXYWDQDPREBGAAPSTFKGTPTENPEYILGLDVPV 1253
 Db 1197 LTFPGGAAPQHPHPAPSPAFNDLXYWDQDPREBGAAPSTFKGTPTENPEYILGLDVPV 1255

RESULT 12
 AAE20479
 ID AAE20479 standard; Protein, 1255 AA.
 AC AAE20479;
 XX 01-JUL-2002 (first entry)
 DT 01-JUL-2002 (first entry)
 DE Human Her-2/neu protein.
 KM Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
 KM human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 1021..1030
 /note="Naturally processed HLA-B44-restricted epitope"
 XX MO200214503-A2.
 XX 21-FEB-2002.
 PF 14-AUG-2001; 2001MO-US41733.
 XX 14-AUG-2000; 2000US-225152P.
 PR 28-SEP-2000; 2000US-236428P.
 PR 21-FEB-2001; 2001US-270520P.
 XX (CORI-) CORIXA CORP.
 PA Hand-Zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
 PI McNeill PD, Vedvick TS;
 PI WPI: 2002-280758/32.
 DR N-PSDB; AAD32743.
 DR
 XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
 PT prevention and diagnosis of cancer, preferably breast cancer
 PT Disclosure; Page 114-117; 129p; English.
 XX The invention relates to an isolated Her-2/Neu polypeptide composition
 CC effective for eliciting an immune response. The invention is useful for
 CC eliciting an immune response in a patient, where the patient is human
 CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
 CC The composition is useful for the therapy and diagnosis of cancer.
 CC Preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
 CC and other compositions for the diagnosis, prevention and treatment of
 CC human malignancies, for stimulating and/or expanding T cells specific for
 CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
 CC patient. The invention is useful for stimulating a T cell response in a
 CC human patient, as probe or primer for nucleic acid hybridization, to
 CC selectively form duplex molecules with complementary stretches of the
 CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
 CC length gene from a suitable library, and to direct expression of a
 CC polypeptide in appropriate host cells. The composition is useful in
 CC prophylactic or therapeutic applications and for the treatment of cancer,
 CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
 CC associated malignancies. The invention is useful in gene therapy. The
 CC present sequence is human Her-2/neu protein.
 XX
 SQ Sequence 1255 AA;

Best Local Similarity 96.7%; Pred. No. 0;
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;
 QY 1 MELAALCRGGLLALLPRGAATOVCTGTDMKLRPLASETHLDMRLHYOGGQVVOGSL 60
 Db 1 MELAALCRGGLLALLPRGAATOVCTGTDMKLRPLASETHLDMRLHYOGGQVVOGSL 60
 QY 61 ELTYLPTNALSFLDIOIEVOGVYLIAHQVAVPQORIRIVRGTOLEFDNVALVLDNG 120
 Db 61 ELTYLPTNALSFLDIOIEVOGVYLIAHQVAVPQORIRIVRGTOLEFDNVALVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLRELQRLSTEILKGVLIQRNQLCYODTILMKDIFKXNOLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSTEILKGVLIQRNQLCYODTILMKDIFKXNOLA 180
 QY 181 LTLIDNRSRACHPCS PMCKSGRCWSESESDQSLTRTVCAGACARCKPLPTDCHEC 240
 Db 181 LTLIDNRSRACHPCS PMCKSGRCWSESESDQSLTRTVCAGACARCKPLPTDCHEC 240
 QY 241 AAGCTGPKHSIDCLAFHNSGICELHCPALVTYNTDRESMNPREGRTFGASCYTACP 300
 Db 241 AAGCTGPKHSIDCLAFHNSGICELHCPALVTYNTDRESMNPREGRTFGASCYTACP 300
 QY 301 YNYLSTDVGSCTLVCEPLHQEYTAEDGTORCEKSPCARVCYGLCMQYIKANSKFTGTT 360
 Db 301 YNYLSTDVGSCTLVCEPLHQEYTAEDGTORCEKSPCARVCYGLCMQYIKANSKFTGTT 360
 QY 361 ELFPAGCKKIFGSLAFLESPFGDDPASNTAPLOPELOVFEETLEETGLYISAMPDLP 420
 Db 361 ELFPAGCKKIFGSLAFLESPFGDDPASNTAPLOPELOVFEETLEETGLYISAMPDLP 420
 QY 421 DLSVFONLOVIRGRILHNGAYSLLTQIGLISWLGRLSLELGSGLALHNNHLCFVHTV 480
 Db 421 DLSVFONLOVIRGRILHNGAYSLLTQIGLISWLGRLSLELGSGLALHNNHLCFVHTV 480
 QY 481 PMDQLEFRNPHOALLHNPANPEDECVGEGIAHQOLARCGKMPPTQCNCQFLRGQSC 540
 Db 481 PMDQLEFRNPHOALLHNPANPEDECVGEGIAHQOLARCGKMPPTQCNCQFLRGQSC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHAHXPFFCVAR 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHAHXPFFCVAR 600
 QY 601 PSQVAPDLSYMPIMKFPDEEGACQCPINCHSCYVDLDDKCPABQRPASPLTSYSAVVG 660
 Db 601 PSQVAPDLSYMPIMKFPDEEGACQCPINCHSCYVDLDDKCPABQRPASPLTSYSAVVG 660
 QY 661 ILVVVLGVVFGILLI-----FNNFTVSPFLRVPKVASHLPLTPSGAMPPOAQWRILK 714
 Db 661 ILVVVLGVVFGILLI-----FNNFTVSPFLRVPKVASHLPLTPSGAMPPOAQWRILK 714
 QY 715 ETELKRVKVLGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVWAG 774
 Db 715 ETELKRVKVLGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVWAG 774
 QY 717 ETELKRVKVLGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVWAG 776
 Db 717 ETELKRVKVLGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVWAG 776
 QY 775 VSPPYVSRLLGICLSTVQVLTQMLPYGCLLDHVENGRISQDOLLMMQOIAAGMSVL 834
 Db 775 VSPPYVSRLLGICLSTVQVLTQMLPYGCLLDHVENGRISQDOLLMMQOIAAGMSVL 834
 QY 835 EDVRVLRDLAARNVLYVSPNHVKITDGLARLLDIDETEVNADGKVPKXMALESILR 894
 Db 835 EDVRVLRDLAARNVLYVSPNHVKITDGLARLLDIDETEVNADGKVPKXMALESILR 894
 QY 896 REFTHOSDVWSGVWVWELMFGAKPYGCIIPAREIPDLLEKERRLPQPPICITDVMYIMV 954
 Db 896 REFTHOSDVWSGVWVWELMFGAKPYGCIIPAREIPDLLEKERRLPQPPICITDVMYIMV 954
 QY 955 KCMWIDSECRPRFRELIVSEFMRADPORFVVIQNEDELGPASPLDSTFYRSILDEDDMGD 1014
 Db 955 KCMWIDSECRPRFRELIVSEFMRADPORFVVIQNEDELGPASPLDSTFYRSILDEDDMGD 1014
 QY 1015 LVDABEYLVPQGGFCPPDPAAGAMVHRRSSSTRGCGGDLTGLPSEBEAPRSLA 1074
 Db 1015 LVDABEYLVPQGGFCPPDPAAGAMVHRRSSSTRGCGGDLTGLPSEBEAPRSLA 1074

Query Match 96.9%; Score 6604; DB 23; Length 1255;

Db 1017 LVDAEEYLVPOQGFCDPAPAGAGWYHHNRSSSTSGGDLTLGLEPSEBAPSPILA 1076
 QY 1075 PSEGAGSDVFDGDLGMAAGAGLGLPTHDSPPLORYSDDPTVPLPSETDGYVAPLTCSPQ 1134
 Db 1077 PSEGAGSDVFDGDLGMAAGAGLGLPTHDSPPLORYSDDPTVPLPSETDGYVAPLTCSPQ 1136
 QY 1135 PEYVNOGDVPPSPREGEPLPARPAGATLERKTLSPKNGVYKDVFPFGAVNPPEV 1194
 Db 1137 PEYVNOGDVPPSPREGEPLPARPAGATLERKTLSPKNGVYKDVFPFGAVNPPEV 1196
 QY 1195 LTPGGAAPQHPPPAPSPAPFNDLYWDQDPPERGAPSTFKGPTAENPEYGLDVPV 1253
 Db 1197 LTPGGAAPQHPPPAPSPAPFNDLYWDQDPPERGAPSTFKGPTAENPEYGLDVPV 1255

RESULT 13

AAM51143
 ID AAM51143 standard; Protein; 1255 AA.

AC AAM51143;

DT 17-JUN-2002 (first entry)

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

KM Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..653 /note= "extracellular domain"

FT Domain 676..1255 /note= "intracellular domain"

FT Domain 990..1255 /note= "phosphorylation domain"

XX MO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIYA CORP.

XX (SMIK-) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX PT or enhancing an immune response to the protein, has Her-2/neu

XX PT extracellular domain fused to Her-2/neu intracellular or

XX PT phosphorylation domain -

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

XX or c-erbB2), an oncogenic self-protein and target for anti-cancer

XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

XX in a variety of cancers, including breast, ovarian, colon, lung and

XX prostate cancer. Her-2/neu is a member of the tyrosine kinase

XX family of receptor-like glycoproteins. It comprises an extracellular

XX domain with homology to the epidermal growth factor receptor

XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal

XX intracellular domain that also shows homology to EGFR. Its

XX overexpression correlates with a poor prognosis in breast and

XX ovarian cancers. The invention provides Her-2/neu fusion

XX proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In

CC preferred fusion proteins, the extracellular domain of a Her-2/neu

CC protein is fused to a Her-2/neu intracellular domain or

CC phosphorylation domain (or its DeltaCD fragment). An immune

CC response to Her-2/neu protein is elicited or enhanced by

CC administering the fusion protein in the form of a vaccine, or by

CC transfecting cells of an animal *ex vivo* with a nucleic acid

CC encoding the fusion protein, and delivering the transfected cells

CC to the animal. The fusion proteins, nucleic acids, and isolated

CC specific T-cells are useful for inhibiting the development of a

CC cancer, especially breast, ovarian, colon, lung or prostate cancer

CC in a patient. T cells that specifically react with a Her-2/neu

CC fusion protein can be used to remove tumour cells from a sample in

CC order to inhibit the development of cancer in a patient.

SQ Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 23; Length 1255;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAATVCTGDMKLRPASBETHLMDLRHLVGGCVVQGL 60
 Db 1 MELAALCRWGLLALLPPGAATVCTGDMKLRPASBETHLMDLRHLVGGCVVQGL 60
 QY 61 ETTYLPTNASTFLDIOEVQGVYLAHNOYRQVLEQLRIYRGTLFEDNYALAVLDNG 120
 Db 61 ETTYLPTNASTFLDIOEVQGVYLAHNOYRQVLEQLRIYRGTLFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVGTASPGGLRELRLSLTEILKGVLIQRNPOLCYQTTIMKDFHKNOLA 180
 Db 121 DPLNNTPTVGTASPGGLRELRLSLTEILKGVLIQRNPOLCYQTTIMKDFHKNOLA 180
 QY 121 DPLNNTPTVGTASPGGLRELRLSLTEILKGVLIQRNPOLCYQTTIMKDFHKNOLA 180
 Db 121 DPLNNTPTVGTASPGGLRELRLSLTEILKGVLIQRNPOLCYQTTIMKDFHKNOLA 180
 QY 181 LTLIDPNNRACHPCSPCKSGRCGSESDCQSLTRYACAGCARCKGPLETDCHEOC 240
 Db 181 LTLIDPNNRACHPCSPCKSGRCGSESDCQSLTRYACAGCARCKGPLETDCHEOC 240
 QY 181 LTLIDPNNRACHPCSPCKSGRCGSESDCQSLTRYACAGCARCKGPLETDCHEOC 240
 Db 181 LTLIDPNNRACHPCSPCKSGRCGSESDCQSLTRYACAGCARCKGPLETDCHEOC 240
 QY 241 AAGCTGPKASDCLACLFHNSGICELACBALVYNTDTFESMPNEGRVTFGASCVTAAP 300
 Db 241 AAGCTGPKASDCLACLFHNSGICELACBALVYNTDTFESMPNEGRVTFGASCVTAAP 300
 QY 241 AAGCTGPKASDCLACLFHNSGICELACBALVYNTDTFESMPNEGRVTFGASCVTAAP 300
 Db 241 AAGCTGPKASDCLACLFHNSGICELACBALVYNTDTFESMPNEGRVTFGASCVTAAP 300
 QY 301 YNYLSTDVSSCTLVCPHNOEYTAEDGQRCCKSPCARVYGLGMQVIXANSKFIGIT 360
 Db 301 YNYLSTDVSSCTLVCPHNOEYTAEDGQRCCKSPCARVYGLGMQVIXANSKFIGIT 360
 QY 301 YNYLSTDVSSCTLVCPHNOEYTAEDGQRCCKSPCARVYGLGMQVIXANSKFIGIT 360
 Db 301 YNYLSTDVSSCTLVCPHNOEYTAEDGQRCCKSPCARVYGLGMQVIXANSKFIGIT 360
 QY 361 ELFEAGCKKIFGSLAFLEPSFDGDPASNTAPLOPELOVFEETLEBITGYLYISAMPDSL 420
 Db 361 ELFEAGCKKIFGSLAFLEPSFDGDPASNTAPLOPELOVFEETLEBITGYLYISAMPDSL 420
 QY 361 ELFEAGCKKIFGSLAFLEPSFDGDPASNTAPLOPELOVFEETLEBITGYLYISAMPDSL 420
 Db 361 ELFEAGCKKIFGSLAFLEPSFDGDPASNTAPLOPELOVFEETLEBITGYLYISAMPDSL 420
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLSRLBELSGIALIHHNHLFCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLSRLBELSGIALIHHNHLFCFVHTV 480
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLSRLBELSGIALIHHNHLFCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLSRLBELSGIALIHHNHLFCFVHTV 480
 QY 481 PWDOLFRRNHQALLTANPEDECVGEGLACHQOLCARGCWCPGTQCVCNSQFLRGQEC 540
 Db 481 PWDOLFRRNHQALLTANPEDECVGEGLACHQOLCARGCWCPGTQCVCNSQFLRGQEC 540
 QY 481 PWDOLFRRNHQALLTANPEDECVGEGLACHQOLCARGCWCPGTQCVCNSQFLRGQEC 540
 Db 481 PWDOLFRRNHQALLTANPEDECVGEGLACHQOLCARGCWCPGTQCVCNSQFLRGQEC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHECPQNSVTCFGEADQCAACHYKDPFCVAPC 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHECPQNSVTCFGEADQCAACHYKDPFCVAPC 600
 QY 541 VEECRVLOGLPREYVNAHCLPCHECPQNSVTCFGEADQCAACHYKDPFCVAPC 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHECPQNSVTCFGEADQCAACHYKDPFCVAPC 600
 QY 601 PSGVPRDLSPYIMKFPDEGACQCPINCHSCVDLDDKGPAPQASPLTISAVVG 660
 Db 601 PSGVPRDLSPYIMKFPDEGACQCPINCHSCVDLDDKGPAPQASPLTISAVVG 660
 QY 601 PSGVPRDLSPYIMKFPDEGACQCPINCHSCVDLDDKGPAPQASPLTISAVVG 660
 Db 601 PSGVPRDLSPYIMKFPDEGACQCPINCHSCVDLDDKGPAPQASPLTISAVVG 660
 QY 661 ILVVVLGVVFGILT-----FNNFTVSFWLEVPVYASHSLEPLTPSGAMPQAOIRILK 716
 Db 661 ILVVVLGVVFGILT-----FNNFTVSFWLEVPVYASHSLEPLTPSGAMPQAOIRILK 716
 QY 715 ETEIRKVKVLSGAGGVYKGIWIPDGENVKIPVAKYLRNTSKAKKEITDEAYVWAG 774
 Db 715 ETEIRKVKVLSGAGGVYKGIWIPDGENVKIPVAKYLRNTSKAKKEITDEAYVWAG 774
 QY 717 ETEIRKVKVLSGAGGVYKGIWIPDGENVKIPVAKYLRNTSKAKKEITDEAYVWAG 776
 Db 717 ETEIRKVKVLSGAGGVYKGIWIPDGENVKIPVAKYLRNTSKAKKEITDEAYVWAG 776

QY 775 VGSFVSRLLIGICLTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 834
 Db 777 VGSFVSRLLIGICLTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 836
 QY 835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESTLR 894
 Db 837 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESTLR 896
 QY 895 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGEELPQPICTIDVYIMV 954
 Db 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGEELPQPICTIDVYIMV 956
 QY 955 KCMWIDSECRPRFRELVSFSSRMARDPQRFVYIIONEDLGPASPLDSTFYRSLLDDMDGD 1014
 Db 957 KCMWIDSECRPRFRELVSFSSRMARDPQRFVYIIONEDLGPASPLDSTFYRSLLDDMDGD 1016
 QY 1015 LVDAEYLVPQGGFFCCDPAPAGAGMYHHRSSSTSGGGDLTLGLEPSEBEAPRSPLA 1074
 Db 1017 LVDAEYLVPQGGFFCCDPAPAGAGMYHHRSSSTSGGGDLTLGLEPSEBEAPRSPLA 1076
 QY 1075 PSEGAGSDVFDGDLGMGAAGKLGSLPTHDPSPLOGYSEDPTVPLPSETDGYVAPLTCSPQ 1134
 Db 1077 PSEGAGSDVFDGDLGMGAAGKLGSLPTHDPSPLOGYSEDPTVPLPSETDGYVAPLTCSPQ 1136
 QY 1135 PEYVNOPDVREPPSPREGPLPARAPAGATLERAKTISPGKNGYKDVAFGAVENPEY 1194
 Db 1137 PEYVNOPDVREPPSPREGPLPARAPAGATLERAKTISPGKNGYKDVAFGAVENPEY 1196
 QY 1195 LTPOGGAAPQPHPPAPSPAFDNLYYWDQPPERKAPSTFKGTPTAENPEYLGLDVVPV 1253
 Db 1197 LTPOGGAAPQPHPPAPSPAFDNLYYWDQPPERKAPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 14
 ID AAU77114 standard; Protein: 1255 AA.
 AC AAU77114:
 DT 05-JUN-2002 (first entry)
 XX Human Her-2/neu polypeptide.
 XX Human Her-2/neu polypeptide.
 KM Human; Her-2/neu; cytostatic; haematological malignancy; CML;
 KM acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
 KM chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
 KM Hodgkin's lymphoma; T cell therapy.
 XX Homo sapiens.
 OS Homo sapiens.
 XX WO200213847-A2.
 PN 21-FEB-2002.
 PD 13-AUG-2001; 2001WO-US25408.
 PF 14-AUG-2000; 2000US-0638280.
 PR 28-SEP-2000; 2000US-0675904.
 XX (COI-1) CORIXA CORP.
 PA Gaiger A, Cheever MA, Hand-zimmermann S;
 PI WPI; 2002-280741/32.
 DR N-PSDB; ABX10730.
 XX Inhibiting haematological malignancy development by administering
 PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
 PT encoding the polypeptide, or antigen presenting cells expressing the
 PT polypeptide
 PS Disclosure; Page 71-74; 74pp; English.
 XX

CC The invention relates to a method for inhibiting development of
 CC haematological malignancy in a patient by administering a polypeptide
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
 CC encoding the polypeptide. Antigen presenting cells that express the
 CC protein can also be administered. The sequences are used for inhibiting
 CC development of haematological malignancy such as acute myelogenous
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
 XX

Sequence 1255 AA;
 Query Match 96.9%; Score 6604; DB 23; Length 1255;
 Best Local Similarity 96.7%; Pred. No. 0;
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 METALCRGGLLALLPFGAASVQVGTGDMKRLPASETHLDMRLHYGGGVQGNL 60
 Db 1 METALCRGGLLALLPFGAASVQVGTGDMKRLPASETHLDMRLHYGGGVQGNL 60
 QY 61 ELTYLPTNASTSLQIQEVQGVYLAHNVQVPPQRLRYRGTLQFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASTSLQIQEVQGVYLAHNVQVPPQRLRYRGTLQFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVTGASPGGLREJLRLSTELIKGVYLIORNPQLCYDITLWKDIFHKNNQLA 180
 Db 121 DPLNNTPTVTGASPGGLREJLRLSTELIKGVYLIORNPQLCYDITLWKDIFHKNNQLA 180
 QY 181 LTIIDNRBRACHPCSPMKSGRCWSESESDQSLTRYACAGGACRCRPLPTDCHEQC 240
 Db 181 LTIIDNRBRACHPCSPMKSGRCWSESESDQSLTRYACAGGACRCRPLPTDCHEQC 240
 QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDRESMPNEGRYTFGASCTYACP 300
 Db 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDRESMPNEGRYTFGASCTYACP 300
 QY 301 YNYLSTDVSGCTLVCPHNOEVTABEDGTGRCCKSPKRCVYCYGLMOYIKANSKFIGIT 360
 Db 301 YNYLSTDVSGCTLVCPHNOEVTABEDGTGRCCKSPKRCVYCYGLMEHLREVAVTSAN 360
 QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPQPELOVFEFLTEITGLYISAMPDSL 420
 Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPQPELOVFEFLTEITGLYISAMPDSL 420
 QY 421 DLSVFONLQVIRGRIHNGAYSLTQIGLISWGLRSLRELSGLALIHNNHLCFVHTV 480
 Db 421 DLSVFONLQVIRGRIHNGAYSLTQIGLISWGLRSLRELSGLALIHNNHLCFVHTV 480
 QY 481 PMDOLFRNPHOALLHTANPEDECVGEGLACHQLCARHCWGPPTQCVNCSQPLRGQEC 540
 Db 481 PMDOLFRNPHOALLHTANPEDECVGEGLACHQLCARHCWGPPTQCVNCSQPLRGQEC 540
 QY 541 VEECRVLOGLPREYVNAHRCLECHPEQOPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 Db 541 VEECRVLOGLPREYVNAHRCLECHPEQOPONGSVTCFGEADQCVACAHYKDPFCVARC 600
 QY 601 PSQVYPDLSYMPIMKPEDEGACOPCPINCHTSCVDLDDKGPAPQASPLTSISAVVG 660
 Db 601 PSQVYPDLSYMPIMKPEDEGACOPCPINCHTSCVDLDDKGPAPQASPLTSISAVVG 660
 QY 661 ILVVYVLGVVFGILI-----FNFVFSFWLRVPKVASHLEPLTPSGAMPQAOAMRLIK 714
 Db 661 ILVVYVLGVVFGILIKRQOKIRKYTMRLRLQETEL-----VEPLTPSGAMPQAOAMRLIK 716
 QY 715 ETELKRVKVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKETILDEAYVYAG 774
 Db 715 ETELKRVKVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKETILDEAYVYAG 776
 QY 775 VGSFVSRLLIGICLTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 834
 Db 775 VGSFVSRLLIGICLTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 836
 QY 835 EDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESTLR 894

Qy 1075 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQ 1134
Db 1077 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQ 1136
Qy 1135 PEYVNPDPVRPOPSPREGLPAPAPAGATLERAKTLSPGKNGVXKDVFAFGAVENPEY 1194
Db 1137 PEYVNPDPVRPOPSPREGLPAPAPAGATLERAKTLSPGKNGVXKDVFAFGAVENPEY 1196
Qy 1195 LTPQGGAAPQHPHPPAFSPAFDNLXYWDQDPPEGAPSPSTFKGTPTAENPEYLGLDVPV 1253
Db 1197 LTPQGGAAPQHPHPPAFSPAFDNLXYWDQDPPEGAPSPSTFKGTPTAENPEYLGLDVPV 1255

Search completed: July 22, 2003, 09:16:38
Job time : 41.875 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.3921 Seconds

(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-695-709-14

Perfect score: 6853

Sequence: 1 MELAALCPWGLLALPFGA.....TEKGTAEPEYLGVDPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6641	96.9	1255	1 A24571	protein-tyrosine k
2	5844	85.3	1260	1 TVRINU	protein-tyrosine k
3	5834.5	85.1	1254	2 T48161	p-185 precursor
4	3065	44.7	1210	1 GQHDE	epidermal growth f
5	3036	44.3	1210	2 A53183	epidermal growth f
6	3013.5	44.0	1223	1 TVCHLV	epidermal growth f
7	2884.5	42.1	1308	2 A47253	epidermal growth f
8	2592	37.8	1366	1 S06142	protein-tyrosine k
9	2334.5	34.1	1342	2 A36223	kinase-related tra
10	2248.5	32.8	1359	1 UC4387	epidermal growth f
11	1685.5	24.6	698	2 TVFVLV	protein-tyrosine k
12	1622	23.7	604	1 TVVUH	protein-tyrosine k
13	1569	22.9	544	2 S35745	protein-tyrosine k
14	1562	22.8	545	2 S00727	kinase-related tra
15	1557.5	22.7	1330	1 GQFPE	epidermal growth f
16	1545	22.5	540	2 B44776	protein-tyrosine k
17	1543	22.0	540	1 TVFVEB	protein-tyrosine k
18	1509	22.0	644	2 A36325	epidermal growth f
19	1241	18.1	1333	2 E88257	protein-tyrosine k
20	1241	18.1	1374	2 S70712	protein-tyrosine k
21	1153	16.8	1369	2 S70713	protein-tyrosine k
22	1129	16.5	1717	1 A45558	epidermal growth f
23	1126	16.4	527	2 A42032	epidermal growth f
24	875.5	14.2	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	700	10.2	1363	2 T43220	insulin-like growth
28	670	9.8	1383	2 A36080	insulin receptor p
29	669.5	9.8	1372	2 A34157	insulin receptor p

30	668	9.7	1382	1 INHR	insulin receptor p
31	655	9.6	1477	2 T18534	protein-tyrosine k
32	653	9.5	1607	2 T43212	insulin-like growth
33	650.5	9.5	1300	2 A36302	insulin receptor p
34	638	9.3	1268	2 B36502	insulin receptor p
35	596	8.7	1367	1 IGHUR1	insulin-like growth
36	590	8.6	1390	2 T30346	insulin receptor p
37	588.5	8.6	2148	1 A56081	insulin receptor p
38	582	8.5	1371	2 S57245	insulin receptor p
39	577	8.4	1371	2 A33837	insulin-like growth
40	574	8.4	1114	1 S05882	protein-tyrosine k
41	571	8.3	987	2 A54092	protein-tyrosine k
42	569.5	8.3	952	2 I50612	protein-tyrosine k
43	566.5	8.3	984	2 A39753	protein-tyrosine k
44	562.5	8.2	977	2 S49004	tyrosine kinase Mp
45	556.5	8.1	976	2 A36355	protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein
C:Species: Homo sapiens (man)
C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622

R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, Nature 319, 230-234, 1986

A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal gro

A:Reference number: A24571; MUID:86118663; PMID:3003577

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 <YAM>

A:Cross-references: GB:X03363; NID:931197; PIDN:CAA27060.1; PID:931198

R:Semba, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/ep

A:Reference number: A25491; MUID:86016729; PMID:2995967

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 <SEM>

A:Cross-references: GB:M1767; NID:9182163; PIDN:AAA35808.1; PID:9553282

R:Gustussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seebu Science 230, 1132-1139, 1985

A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares ch

A:Reference number: A44188; MUID:86070181; PMID:2999974

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 <COU>

A:Cross-references: GB:M12036; NID:9183988; PIDN:AAA35978.1; PID:9183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517, 'RALU', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU>

A:Cross-references: GB:M11730; NID:9183986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985

A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089

A:Accession: I59509

A:Molecule type: DNA

A:Residues: 832-909 <REX>

A:Cross-references: GB:I29395; NID:9459807; PIDN:AAA35809.1; PID:9459808

R:Pal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987

A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcript

A:Reference number: I57622; MUID:87286898; PMID:3039351

A:Accession: I57622

A:Molecule type: DNA

A:Residues: translated from GB/EMBL/DBDT

A:Molecule type: DNA

A:Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:g183989; PIDN:AAA5637.1; PID:g553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetics:
 A:Gene: GDB:ERBB2; NGL; NEU; HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the list of introns is incomplete
 C:Function:
 A:Description: catalyzes the phosphorylation of a peptide tyrosine residue by ATP
 A:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
 F:1-31/Domain: signal sequence #status predicted <SIG>
 F:122-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:122-653/Domain: extracellular #status predicted <EXT>
 F:170-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F:654-675/Domain: transmembrane #status predicted <TM>
 F:676-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Region: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:68,124,187,259,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:753/Active site: Lys #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.9% Score 6641; DB 1; Length 1255;
 Best Local Similarity 97.3% Pred No.5.5e-266;
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

QY 1 MELALCRWGLLALPPGAASVCTGTDMKRLRLPASPEHMLRLHLYGGCCVVGNTL 60
 DB 1 MELALCRWGLLALPPGAASVCTGTDMKRLRLPASPEHMLRLHLYGGCCVVGNTL 60
 QY 61 ELTYLPTNASLFLDIOEGVYVLAHNOVROVLRQRLRVGRGQFEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLFLDIOEGVYVLAHNOVROVLRQRLRVGRGQFEDNYALAVLDNG 120
 QY 121 DPLNNTVTGASPGGLRELOLRSLTEILKGVLIQORNPOLCYOQDILMKDIFKNNOLA 180
 DB 121 DPLNNTVTGASPGGLRELOLRSLTEILKGVLIQORNPOLCYOQDILMKDIFKNNOLA 180
 QY 181 LTLIDTNSRACHPCSPMCKSRGCESSDDCOSLTRVACGACRCRGLPTCCHEQC 240
 DB 181 LTLIDTNSRACHPCSPMCKSRGCESSDDCOSLTRVACGACRCRGLPTCCHEQC 240
 QY 241 AAGCTGPKHSDCLALCFHNSGICELHCPALVTYNTDTFEEMPNEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLALCFHNSGICELHCPALVTYNTDTFEEMPNEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVPLHNOEYTAADGTQRCCKSKPCARVCYGLMOYITANSKFTGIT 360
 DB 301 YNYLSTDVSGCTLVPLHNOEYTAADGTQRCCKSKPCARVCYGLMOYITANSKFTGIT 360
 QY 361 YNYLSTDVSGCTLVPLHNOEYTAADGTQRCCKSKPCARVCYGLMOYITANSKFTGIT 360
 DB 361 YNYLSTDVSGCTLVPLHNOEYTAADGTQRCCKSKPCARVCYGLMOYITANSKFTGIT 360
 QY 420 ELLEFGCKKIFGSLALFLPESFGDPSANTAPLQPOLQVETLEETLYLISAMPDLSL 420
 DB 420 ELLEFGCKKIFGSLALFLPESFGDPSANTAPLQPOLQVETLEETLYLISAMPDLSL 420
 QY 480 DLSVFQNTQVIRGRILHNGAYSLTQGLGSLMGLRSRLREISGSLALHNTHTLCFVTV 480
 DB 480 DLSVFQNTQVIRGRILHNGAYSLTQGLGSLMGLRSRLREISGSLALHNTHTLCFVTV 480
 QY 540 PMDQLFRNPHQALLTARPEDECVGEGLAGHQCACARHCHGPGGTQCVNCSQFLRGEC 540
 DB 540 PMDQLFRNPHQALLTARPEDECVGEGLAGHQCACARHCHGPGGTQCVNCSQFLRGEC 540
 QY 600 VEEGRVLOGLPREYVNAHCLPCHDECOQPNQSVTCFSPADQCVACAHYDPPFCVARC 600
 DB 600 VEEGRVLOGLPREYVNAHCLPCHDECOQPNQSVTCFSPADQCVACAHYDPPFCVARC 600
 QY 660 PSGVPRDLSYMPIMKFPDEEGACQCPINCHTSCVDLDDKGCAPABQASPLTISAVNG 660
 DB 660 PSGVPRDLSYMPIMKFPDEEGACQCPINCHTSCVDLDDKGCAPABQASPLTISAVNG 660

DB 601 PSGVPRDLSYMPIMKFPDEEGACQCPINCHTSCVDLDDKGCAPABQASPLTISAVNG 660
 QY 661 ILVVVLGVVFGILLIKROOKIRKTYMRLLLOETELVEVLTSSGAMPNQAQRILKEFNK 720
 DB 661 ILVVVLGVVFGILLIKROOKIRKTYMRLLLOETELVEVLTSSGAMPNQAQRILKEFNK 720
 QY 721 FTYSFMLRVKVSAS-HLETVYKGIWIPGENVKIPVAIKVIRENTSPKANEILDEAVY 779
 DB 720 -----LRKVKVLSGAFGTVYKGIWIPGENVKIPVAIKVIRENTSPKANEILDEAVY 773
 QY 780 MAGVSPYVSRLLIGLICTSTVQLVQLMEYGLLDHVRNRRGLSQDILLNMCQIAKGM 839
 DB 774 MAGVSPYVSRLLIGLICTSTVQLVQLMEYGLLDHVRNRRGLSQDILLNMCQIAKGM 833
 QY 840 SYLEDVRLVHRDLAANVVKSPNHVKINDPGLARLLDIDETEVHADGGKVPKMMALS 899
 DB 834 SYLEDVRLVHRDLAANVVKSPNHVKINDPGLARLLDIDETEVHADGGKVPKMMALS 893
 QY 900 ILRRPTTHSDVMSYGVTVWELMTFGAKPYDGI-PAREIDLLEKGERLPQPICTIDVYM 959
 DB 894 ILRRPTTHSDVMSYGVTVWELMTFGAKPYDGI-PAREIDLLEKGERLPQPICTIDVYM 953
 QY 960 IMVKCMYIDSECPFRRELVSERSRARDPQRFVYQNDLGPASFLDSTFRSLLED 1019
 DB 954 IMVKCMYIDSECPFRRELVSERSRARDPQRFVYQNDLGPASFLDSTFRSLLED 1013
 QY 1020 MGDLVAEEFLVPOQGFCEPAPAGAGVHHRSSSTRSGGDLTLGLEPSEBAPRS 1079
 DB 1014 MGDLVAEEFLVPOQGFCEPAPAGAGVHHRSSSTRSGGDLTLGLEPSEBAPRS 1073
 QY 1080 PLAPSGAGSDVFDGLGMAKAGLQSLPTHDSPLQRYSEDPVFLPSETDGYVAPLTC 1139
 DB 1074 PLAPSGAGSDVFDGLGMAKAGLQSLPTHDSPLQRYSEDPVFLPSETDGYVAPLTC 1133
 QY 1140 SPQEVYNQPDVAPQPPSPRECEPLPAAPAGATLEPAKTLSPKSNVVDVPAFGAVEN 1199
 DB 1134 SPQEVYNQPDVAPQPPSPRECEPLPAAPAGATLEPAKTLSPKSNVVDVPAFGAVEN 1193
 QY 1200 PEYLTPQGAAPQHPPEPAPFAFDNLVYWDQPPRPGAPPTFKGTPTAENBEVGLDV 1259
 DB 1194 PEYLTPQGAAPQHPPEPAPFAFDNLVYWDQPPRPGAPPTFKGTPTAENBEVGLDV 1253
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 2
 TIRTNV
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 31-Dec-1988 #sequence: revision 31-Dec-1988 #text: change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Bartmann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related prote
 A:Reference number: A24562; M01D:86118662; PMID:3945311
 A:Accession: A24562
 A:Molecule type: mRNA
 A:Residues: 1-1260 <BAR>
 A:Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
 R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; C
 Carlingense 12, 1975-1978, 1991
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals
 2-thiazolyl[1]formamide or N-methyl-N-nitrosourea.
 A:Reference number: A61204; M01D:92035293; PMID:1682063
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663, 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 C:Genetics:
 A:Gene: neu

C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:1-13/Domain: signal sequence #status predicted <Sig>
 F:120-1260/Product: protein-tyrosine kinase neu #status predicted <Mat>
 F:658-680/Domain: transmembrane #status predicted <TMN>
 F:723-988/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:721.191.263.535.576.634/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted
 F:758/Active site: lys #status predicted
 F:1882.1227.1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.3%; Score 5844; DB 1; Length 1260;
 Best Local Similarity 85.7%; Pred. No. 3.2e-233;
 Matches 1083; Conservative 52; Mismatches 119; Indels 10; Gaps 4;

QY 1 MELALCRWGLLALLPPGAASVQVCTGTDKRLPLPASPETHLDMRLHYGCGVYQGNL 60
 DB 4 MELAAWCRWGLLALLPPGIAGVQVCTGTDKRLPLPASPETHLDMRLHYGCGVYQGNL 63
 QY 61 ELTYLPTNASLFLDIOEVOGYVLIHANQVQVPLQRLIRVGTQLFEDNYALAVDNG 120
 DB 64 ELTYLPANASLSTLQIOEVOGYVLIHANQVQVPLQRLIRVGTQLFEDNYALAVDNR 123
 QY 121 DPLNNTTAVT-GASPGGLREQLRS:TEILKGVLIQRNPOLCYODTILMKDIFKKNQL 179
 DB 124 DPDQNVAASTPGTPEGLREQLRS:TEILKGVLIQRNPOLCYODTILMKDIFKKNQL 183
 QY 180 ALTLIDTNSRACHPCSPMKGSRGSESEDCQS:TRIVCAGGCRCKGRLPTCCHEQ 239
 DB 184 APVDITNSRACHPCSPMKGSRGSESEDCQS:TRIVCAGGCRCKGRLPTCCHEQ 243
 QY 240 CAAGTGPXHSDDLACIAFNHSGICELCPALVTYNTDTFESMPNDEGRYTGASCVTA 299
 DB 244 CAAGTGPXHSDDLACIAFNHSGICELCPALVTYNTDTFESMPNDEGRYTGASCVTA 303
 QY 300 PNYVISTDVSGCTVCEPLNQEYTAEDGQRCCKSKPCARVYGIQMGYIYANKKFI 359
 DB 304 PNYVISTDVSGCTVCEPLNQEYTAEDGQRCCKSKPCARVYGIQMGYIYANKKFI 363
 QY 360 TELEFAGCKKIFGSLAFLPESPDGASNTAPLOPEQLQVFELEETIGYLIASMPDL 419
 DB 364 NYQEPDGGCKKIFGSLAFLPESPDGASNTAPLOPEQLQVFELEETIGYLIASMPDL 423
 QY 420 PDLVSQNLQVIRGRILHNGASVLTQIGISWLGIRSLRELSGIALIHNHTILCFVHT 479
 DB 424 RDLVSQNLRIIRGRILHNGASVLTQIGISWLGIRSLRELSGIALIHNHTILCFVHT 483
 QY 480 VPMDQLFRNPHQALLTANRPEDE-CVGEGLACHQLCARAGCMGPGPTQCVNCSQFLRG 538
 DB 484 VPMDQLFRNPHQALLHSGNRPEDELCVSSGIVNNSICAHQCMGPGPTQCVNCSHFLRG 543
 QY 539 ECVEECRVLIQGLPREVYNAHCLPCHEPCQPNQSVTGCPEPADQCVACAHYKDEPCVA 598
 DB 544 ECVEECRVKGLPREVYSDKRLPCHEPCQPNQSVTGCPEPADQCVACAHYKDEPCVA 603
 QY 599 RCPSSGVKPDLSYMPKPPDEBGAQCPINCHSGVDDDDGCPAPKORASPLTSIVSAV 658
 DB 604 RCPSSGVKPDLSYMPKPPDEBGAQCPINCHSGVDDDDGCPAPKORASPLTSIVSAV 663
 QY 659 VGLILVAVLVGFGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOQOMRLKEF 718
 DB 664 EGVLLFLILVAVLVGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOQOMRLKEF 723
 QY 719 NNFTVAFMLRVPVYAS-SHLETYKGIWT PDGENYKI PAIKVLRNTPSPKANKELIDA 777
 DB 724 E-----LRKVVLVGSAFGTYKGIW PDGENYKI PAIKVLRNTPSPKANKELIDA 776
 QY 778 YVNAVGSPYVSLIGICLTSVQVLTQIMPYGCLLDHVRNKGSLGODLNNMCQIAK 837
 DB 777 YVNAVGSPYVSLIGICLTSVQVLTQIMPYGCLLDHVRNKGSLGODLNNMCQIAK 836
 QY 838 GMSYLEDAVLNHRDLAARNVLYKSPNHVITIDFGRLALDIDETEHADGKVPKIMMAL 897

DB 837 GMSYLEDAVLNHRDLAARNVLYKSPNHVITIDFGRLALDIDETEHADGKVPKIMMAL 896
 QY 898 ESILRRRFTHOSDVWSGVYTWELMTGAKPYDGI PARBEIPDLLEKGERLPQPICTIDV 957
 DB 897 ESILRRRFTHOSDVWSGVYTWELMTGAKPYDGI PARBEIPDLLEKGERLPQPICTIDV 956
 QY 958 YMIWVKCMIDSECRPPRELVEFSPPARDPQRFVYI QNEDLPSSGMDSTFRSLLED 1017
 DB 957 YMIWVKCMIDSECRPPRELVEFSPPARDPQRFVYI QNEDLPSSGMDSTFRSLLED 1016
 QY 1018 DDMGDVDAEEYLVPOGFFCPDPAAGQVNHRRHSSSTRSGGDTITLGPSEEDAP 1077
 DB 1017 DDMGDVDAEEYLVPOGFFCPDPAAGQVNHRRHSSSTRSGGDTITLGPSEEDAP 1076
 QY 1078 RSPILAPSGASDVFDGDLGCAKGIQSLP:THPSPLQXSEPTPLPSETGYVAPL 1137
 DB 1077 RSPILAPSGASDVFDGDLGCAKGIQSLP:THPSPLQXSEPTPLPSETGYVAPL 1136
 QY 1138 TCSPPQPEYVNPQVPRPQPSPREGRLPAARPAGATLERAKTILSPGKGVVDVAFGAV 1197
 DB 1137 ACSPPQPEYVNPQVPRPQPSPREGRLPAARPAGATLERAKTILSPGKGVVDVAFGAV 1196
 QY 1198 ENPEYLVREGTASBP:PHSPAFSPAFDNLVYWDQDPERRCAPPTFKTPTAENPEYGL 1257
 DB 1197 ENPEYLVREGTASBP:PHSPAFSPAFDNLVYWDQDPERRCAPPTFKTPTAENPEYGL 1256
 QY 1258 DVPV 1261
 DB 1257 DVPV 1260

RESULT 3

148161
 P-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 R:Nakamura, T.; Uehijima, T.; Ishikawa, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishi
 Gene 140, 251-255, 1994
 A>Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Reference number: 148161; MUID:94193007; PMID:7908275
 A:Accession: 148161
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1254 <RES>
 A:Cross-references: GB:D16295; NID:G493236; PID:BA03801.1; PID:G747595
 C:Genetics:
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP
 F:718-983/Domain: protein kinase ATP-binding motif
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.1%; Score 5834.5; DB 2; Length 1254;
 Best Local Similarity 85.3%; Pred. No. 7.8e-233;
 Matches 1077; Conservative 61; Mismatches 115; Indels 9; Gaps 3;

QY 1 MELALCRWGLLALLPPGAASVQVCTGTDKRLPLPASPETHLDMRLHYGCGVYQGNL 60
 DB 1 MELAAWCRWGLLALLSPGASVQVCTGTDKRLPLPASPETHLDMRLHYGCGVYQGNL 60
 QY 61 ELTYLPTNASLFLDIOEVOGYVLIHANQVQVPLQRLIRVGTQLFEDNYALAVDNG 120
 DB 61 ELTYLPANASLSTLQIOEVOGYVLIHANQVQVPLQRLIRVGTQLFEDNYALAVDNR 120
 QY 121 DPLNNTTAVT-GASPGGLREQLRS:TEILKGVLIQRNPOLCYODTILMKDIFKKNQLA 180
 DB 121 DPLNNTTAVTGRTEGRLREQLRS:TEILKGVLIQRNPOLCYODTILMKDIFKKNQLA 180
 QY 181 LTLIDTNSRACHPCSPMKGSRGSESEDCQS:TRIVCAGGCRCKGRLPTCCHEQ 240
 DB 181 PVDITNSRACHPCSPMKGSRGSESEDCQS:TRIVCAGGCRCKGRLPTCCHEQ 240

QY 241 AAGCTGPKXSDCLACIHFNNHSGICEIHCAPALVTYNTDTESMNPBEGRTYTFGASCTTAC 300
 DB 241 AAGCTGPKXSDCLACIHFNNHSGICEIHCAPALVTYNTDTESMNPBEGRTYTFGASCTTAC 300
 QY 301 YNLTSTDVSGCTVPCPLHNOEVAEDGTORCEKSPKACVCGLMQVYIKANSKFIG:IT 360
 DB 301 YNLTSTDVSGCTVPCPLHNOEVAEDGTORCEKSPKACVCGLMQVYIKANSKFIG:IT 360
 QY 361 ELDPAGCKI FGSGLAPESFDPPASNTAPLOPEOLQVETLEETITGVLISAMPDLSIP 420
 DB 361 ELDPAGCKI FGSGLAPESFDPPASNTAPLOPEOLQVETLEETITGVLISAMPDLSIP 420
 QY 421 DLSPQNLQVIRRIILHNGAYSLTLOOLGISWLGSLRLSGLALHNNHNLGVHNV 480
 DB 421 DLSPQNLQVIRRIILHNGAYSLTLOOLGISWLGSLRLSGLALHNNHNLGVHNV 480
 QY 481 PMDQLEPNPHQALLHTANRDEDECVBEGGLAQHQLCARHGWGPGPTQVNCGQFLRGQEC 540
 DB 481 PMDQLEPNPHQALLHTANRDEDECVBEGGLAQHQLCARHGWGPGPTQVNCGQFLRGQEC 540
 QY 541 VECRCVLOGLPREYVNAHCLPCHPEGQPNQSVTQEGPADQCVCAHYKDPFCVAC 600
 DB 541 VECRCVLOGLPREYVNAHCLPCHPEGQPNQSVTQEGPADQCVCAHYKDPFCVAC 600
 QY 601 PSGVKKPDLSTYMPIMKPPDEBGAQCPINCTHSCVDLDDKCEBQASPLTSISAVVG 660
 DB 601 PSGVKKPDLSTYMPIMKPPDEBGAQCPINCTHSCVDLDDKCEBQASPLTSISAVVG 660
 QY 661 ILVLVGLVVGFLIKRROOKIKRYTMRLLQETELVEPLTSGAMPNQAQRIKEPN 720
 DB 661 ILVLVGLVVGFLIKRROOKIKRYTMRLLQETELVEPLTSGAMPNQAQRIKEPN 720
 QY 721 FTVSFWLVRVKSAS-HLETVYKGIWIPDGENVKIPVALKYLENTPSPANKELDEAV 779
 DB 721 FTVSFWLVRVKSAS-HLETVYKGIWIPDGENVKIPVALKYLENTPSPANKELDEAV 779
 QY 780 MAGVSPYVSRLLIGTICITSTVOLYQTMPIGCLLDHYRENRGLSGQDLINMCQIAKQM 839
 DB 780 MAGVSPYVSRLLIGTICITSTVOLYQTMPIGCLLDHYRENRGLSGQDLINMCQIAKQM 839
 QY 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETVYHADGKVPKIALES 893
 DB 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETVYHADGKVPKIALES 893
 QY 900 ILRRFTHGSDVMSYGVTTWELMTFGAKPIYDGI:PARBIPDLEKGERLPOPICTIDVYM 959
 DB 900 ILRRFTHGSDVMSYGVTTWELMTFGAKPIYDGI:PARBIPDLEKGERLPOPICTIDVYM 959
 QY 960 IMYKCMWIDSECRPFRELSEFSRMAKDPORFVJONEDLGPASPLDSFYRSLLEDD 1019
 DB 960 IMYKCMWIDSECRPFRELSEFSRMAKDPORFVJONEDLGPASPLDSFYRSLLEDD 1019
 QY 1020 MGDVLDAEYLVPOQGFCEPDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEBAPRS 1079
 DB 1020 MGDVLDAEYLVPOQGFCEPDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEBAPRS 1079
 QY 1080 ELAPBEGAGSDVDGDLGMAAGLGLSLPTHDSPLQRYSEDETVLSPSTQVYAPLTC 1139
 DB 1080 ELAPBEGAGSDVDGDLGMAAGLGLSLPTHDSPLQRYSEDETVLSPSTQVYAPLTC 1139
 QY 1140 SPQEVYNQDVPAPQPSPREGLPAPAPAGATLEBAKTLSPKNGVAVDVPFGAVEN 1199
 DB 1140 SPQEVYNQDVPAPQPSPREGLPAPAPAGATLEBAKTLSPKNGVAVDVPFGAVEN 1199
 QY 1200 PEYLTPOGGAPOQHPPAPSPAFDNLVYWDODPERRGAPSPFKGTPTANENYGLDV 1259
 DB 1200 PEYLTPOGGAPOQHPPAPSPAFDNLVYWDODPERRGAPSPFKGTPTANENYGLDV 1259
 QY 1260 PV 1261
 DB 1260 PV 1261
 QY 1253 PV 1254
 DB 1253 PV 1254

RESULT 4

GQ4UE
 epidermal growth factor receptor precursor - human
 N/Contexts: protein-tyrosine kinase (BC 2.7.1.112) erbB
 C/Species: Homo sapiens (man)
 C/date: 15-Nov-1994 #sequence revision 27-Nov-1995 #text change 11-Jun-1999
 C/accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143
 R/Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee,
 rg, P.H.
 Nature 309, 418-425, 1984
 A/Title: Human epidermal growth factor receptor cDNA sequence and aberrant expressi
 A/Reference number: A00641; MUID:84219729; PMID:6328312
 A/Accession: A00641
 A/Molecule type: mRNA
 A/Residues: 1-1210 <U>L>
 A/Cross-references: EMBL:X00588; NID:G31113; PIDN:CAA2240.1; PID:G757924
 A/Note: the authors translated the codon AAG for residue 540 as Asn
 R/ishi, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
 A/Title: Characterization and sequence of the promoter region of the human epiderma
 A/Reference number: A25772; MUID:85270438; PMID:2991899
 A/Accession: A25772
 A/Status: translation not shown
 A/Molecule type: DNA
 A/Residues: 1-29 <ISH>
 A/Cross-references: GB:M1234; NID:G181981; PIDN:AAA2370.1; PID:G553272
 R/Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
 Oncogene Res. 1, 375-396, 1987
 A/Title: The human EGF receptor gene: structure of the 110 kb locus and identifiat
 A/Reference number: S30024; MUID:88217333; PMID:3329716
 A/Accession: S30024
 A/Molecule type: DNA
 A/Residues: 1-29 <HA2>
 A/Cross-references: EMBL:X06370; NID:G31118; PIDN:CAA2668.1; PID:G31119
 R/Haley, J.D.; Waterfield, M.D.
 J. Biol. Chem. 266, 1746-1753, 1991
 A/Title: Contributory effects of de Novo transcription and premature transcript ter
 A/Reference number: A38672; MUID:91107677; PMID:1988448
 A/Accession: A38672
 A/Molecule type: DNA
 A/Residues: 1-29 <NAL>
 A/Cross-references: GB:M38425; NID:G181977; PIDN:AAA63171.1; PID:G553271
 R/Xu, Y.; Ishii, S.; Clark, A.U.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.
 Nature 309, 806-810, 1984
 A/Title: Human epidermal growth factor receptor cDNA is homologous to a variety of
 A/Reference number: A00642; MUID:84245835; PMID:6330563
 A/Accession: A00642
 A/Molecule type: mRNA
 A/Residues: RCMARRA, 150-187, 'KSVIQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 30
 ', 798-799, 'TD', 802-811, 'R', 813-942 <XOY>
 A/Experimental source: A431 human carcinoma cells, which have large numbers of EGF
 R/lin, C.R.; Chen, W.S.; Krutger, W.; Stolarky, L.S.; Weber, W.; Evans, R.M.; Verm
 Science 224, 843-848, 1984
 A/Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica
 A/Reference number: A43615; MUID:84196372; PMID:6326261
 A/Accession: A43615
 A/Molecule type: mRNA
 A/Residues: 713-964 <LIN>
 A/Experimental source: epidermoid carcinoma cell line A431
 R/Stimten, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.J
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984
 A/Reference number: A23062; MUID:85046483; PMID:6093780
 A/Accession: A23062
 A/Molecule type: mRNA
 A/Residues: 1028-1210 <SIM>
 R/weber, W.; G11, G.N.; Speiss, J.
 Science 224, 294-297, 1984
 A/Reference number: A05281; MUID:84172183; PMID:6324343
 A/Accession: A05281
 A/Molecule type: protein
 A/Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R.Ruesco, M.W.; Lukasz, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor
 A:Reference number: A60143; PMID:85182650; PMID:2985590
 A:Accession: A60143
 A:Molecule type: protein
 A:Residues: 740-744, 'X', 746-747 <RUS>
 R.Moczko, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase
 A:Reference number: A38023; PMID:84191554; PMID:6325948
 A:Accession: A38023
 A:Contents: annotation; receptor activity
 A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R.Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 A:Title: Functional independence of the epidermal growth factor receptor from a domain involved in ligand binding
 A:Reference number: A33331; PMID:90003233; PMID:2790960
 A:Contents: annotation; internalization signal
 A:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex
 C:Genetics:
 A:Gene: GDB:EGFR
 A:Cross-references: GDB:120610; OMIM:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-110/Domain: signal sequence #status predicted <SIG>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <E2>
 F:390-600/Domain: EGF receptor extracellular domain repeat <E2>
 F:666-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:110-975/Domain: protein kinase homology <KIN>
 F:118-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:745/Active site: Lys #status experimental

Query Match 44.7%; Score 3065; DB 1; Length 1210;
 Best Local Similarity 48.7%; Pred. No. 5,4e-119;
 Matches 620; Conservative 179; Mismatches 359; Indels 114; Gaps 24;

```

QY 11 LLLALLPPGAA--STOVCTGTDMKLLPLPASPETHLMLRLHYGGCCVVOGNETLTLPTN 68
DB 14 LLLALCPASRALEBEKVCCTSKMLQGLGFEDHFLSLQMPNCEVVLGNLEITTVQGN 73
QY 69 ASLSPLDIOIEOVGVYLAHNOYRQVPLQRLIVRGSTDFEDNYALAVLDNGDPLNNTP 128
DB 74 YDSFLKTIQEVAGVYLAINTVERIPLEVLQIRGNMYEENSALAVLSND----- 126
QY 129 VFGASPGSLREILQSLTELKIGVLIQNPOLCYQDITLMMDIFHKNNQALLTILDTNR 188
DB 127 ---ANKTGLKELPWRLOELHGAIRFNNPALCNVESIQMRIVSSDPLSNMSMDFQNH 183
QY 189 SRACHPCSPWCKSRGCMGESSDDQSLITRTVCAGGA-RCKGLPTDCCHCCGCAAGCTGP 247
DB 184 LGSQCKDCPSCPGSCWGAEEBCKLTKITLCAQCCSGRCRGKSPDCCHNCAGACTGP 243
QY 248 KESDCLALCHFNHSGICELHCPALVYNDTPESMNPGRRTFGASCCTTACPRYVLTSD 307
DB 244 RESDCLVCRKFRDEATCKTCTPPLMKNPTTYQMDVNPBGKYSFGATCVKCPRRNVVVD 303
QY 308 VGSCTLVCPILHNOEVAEDGTORCEKSPKCAVCGYLMQVYKANSKFIQTELE-FAG 366
DB 304 HGSCTVACGADSYEM-EDGVKCKCKCEGCRKVCNGIGIERK-DSLSINATNTHFGN 361
QY 367 CKKIFGSLAFLEPSFPGDPASNTAPLQPSQLOVFEETLEITGVLYISAMPDLPDLSTVQ 426
DB 362 CTSISGDHLPLPAFPGSDFTHTPPDPELDIKTKVKEITGILLQAMPENRTIDHAF 421
QY 427 NLOVIRGLIHNGAVSLTLOGIGISWLG-RSLRELSGGLALHNHNLCPVHTVPDOLF 486

```

DB 422 NLEIRGRTKOHGQSLAVSLNITSLRLSLKEISDDGVIIISGNKMLCVANTIMKKLF 481
 QY 487 ENPHALHTANRPREDCEVGGGLACHOCARHGCMGPPPTGVNCSQFLPQGECEBGRV 546
 DB 482 GTSQCKTKIISKRGNSKATGVCHALCSBGCMPBPBPCVSRNYSRGRCEVDCKTL 541
 QY 547 LOGLPREYVNAHCLPCHPECOPOGVSATCEPBAQCAVNAHYKDPFCVAPCSGVKP 606
 DB 542 LEGREPREVENSECIOCHPECLPQAMNITTCGRGPDNCTICQAHYIDGHCVATCPAGWG 601
 QY 607 DLSTVPIKPFDEBACQPCPLNCHSCVDLDDKCPRAEQASPLTISVSAVY--ILL 663
 DB 602 ENNTL-VKTYADAGVCHLCPNCTYGTGPGLECPNPKRIP--SIATGVAGALLLL 658
 QY 664 VVLDGVVFGILIKRQCKIRKYTRRLQSTELVEPLTPSGAMPQAOMRLIKENNFV 723
 DB 659 VVALGIG---LFMRRIHVRKTRLRLLQSEHVEPLTPSGEAPQALLRIKE----- 709
 QY 724 SFMLRVPKVSASHLETYKGIWIPDGENVKIPVAKIVRENTSPKANEIIDEAYVAGV 783
 DB 710 TEFKKIKYLGAGAFGTGVKGLWIPGEKVKIPVAKIKELREATSPKANEIIDEAYVMSV 769
 QY 784 GSPVYSRLTGCTNSTVOLVQMPYGLLDHVENRGLSDODLLNMCQIAKMSYLE 843
 DB 770 DNPVCRLLGLCLTSTVOLITQMPFGCLLYVEHKNNISQVILNMCVQIAKMSYLE 829
 QY 844 DVRLVHRDLAARNVLYKSPNHVKITDFGLARLDIDETRYADGQKVPDKMALESILRR 903
 DB 830 DRLVHRDLAARNVLYKSPNHVKITDFGLARLDIDETRYADGQKVPDKMALESILRR 889
 QY 904 RFTHSDVWSIGVYWEIMLTGAKPYDGIIPAREIPDLLEKGBRLPQPICTIDVYMIWK 963
 DB 890 IYTHOSDWSIGVYWEIMLTGSKPYDGIIPASEIISILEKEBRLPQPICTIDVYMIWK 949
 QY 964 CMWISSECRPRELSEFSSMARDPQRFVYIQ-NEDLGPASPLDSTFTYRLLEDDMDGD 1022
 DB 950 CMWIDADSRPKFREILIFFSMGARDPQYLYIQTGDERHNLSPDTSNRYRLMDEEDMD 1009
 QY 1023 LYDAEYLVPOQGFPCPDPAQAGMVHRRSSSTRSGGDLTLGLEFSEEBARSPVA 1082
 DB 1010 VYDAEYLVPOQGFPCPDPAQAGMVHRRSSSTRSGGDLTLGLEFSEEBARSPVA 1035
 QY 1083 PSEGAGSVDFDGDGMGAAGLQSLPTDPSPLQRYSDPTVPLPSET--GGVYAPFLTCS 1140
 DB 1036 SLSATSN--NSTVACIDRNGLQSCPIKEDSFLQRYSDPTVPLPSET--GGVYAPFLTCS 1087
 QY 1141 PPEYVNPQDYRPPSPREBPPLPARPAGATLEBRKATLSFGKNGVVDVFAFGAVENP 1200
 DB 1088 PVPEYINQ-SVKKPRAGSVQNPVYHNGPLN-----APSRDPHYQD--PHSTAVGNP 1136
 QY 1201 EYL-TPQGAARQPHPPAPSPARDNLVYNDQ-----DP-----PERGAPSTF 1243
 DB 1137 EYLVNQ-----PTCVNSTFSDSPAHMAQKSHQISLDNPYQODFFPKAKPNIGF 1187
 QY 1244 KGTPTAENPEYL 1255
 DB 1188 KGS-TAENAEYL 1198

RESULT 5
 A53183
 epidermal growth factor receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C:Accession: A53183; A43818; S24944; A28941; S45325; I49643
 R:Unetctete, N.C.; Philippi, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N
 Genes Dev. 8, 359-413, 1994
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
 A:Reference number: A53183; PMID:94170986; PMID:8152525
 A:Accession: A53183
 A:Molecule type: mRNA
 A:Residues: 1-1210 <LUE>
 A:Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
 Oncogene 6, 673-676, 1991
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
 A:Reference number: A43818, MUID:91232866; PMID:2030916
 A:Accession: A43818
 A:Molecule type: mRNA
 A:Residues: 1-714 <AVI>
 A:Cross-references: GB:X59698
 R:Bislinger, D.P.; Serrero, G.
 Submitted to the EMBL Data Library, June 1992
 A:Reference number: S24942
 A:Accession: S24942
 A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heisermann, G.J.; Gili, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941, MUID:8830814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
 R:Hubbs, M.L.; Dunn, A.R.; Alexander, W.S.
 Submitted to the EMBL Data Library, April 1994
 A:Reference number: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CA55587.1; PID:G488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
 A:Reference number: I49643; MUID:93126380; PMID:7678348
 A:Accession: I49643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:I06864; NID:G193001; PIDN:AA53029.1; PID:G567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:770-128/Region: protein kinase ATP-binding motif
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.3%; Score 3036; DB 2; Length 1210;
 Best Local Similarity 48.6%; Pred. No. 8, 4e-118; Indels 118; Gaps 26;
 Matches 622; Conservative 172; Mismatches 367;

244 CTGPHSDCLALPHNSIGCELCPLALVYNTDTERSMNPERRTFGASCVTACPPNY 303
 240 CTGPHSDCLALPHNSIGCELCPLALVYNTDTERSMNPERRTFGASCVTACPPNY 299
 304 LSTDVSGCTLCPLPHNOEVTAEADGTGCEKSKPCARVCGYKANSKFTIGTELE 363
 300 VTTDHGCVARAGDYIEV-EDGIRCKCKDCGCRVKNIGIGERK-DLSLNATNIK 357
 364 -FACCKKIFGSLAFESFDGPPASNTAPLOEOLOVETLEITGHIYISAWPDSL 422
 358 HFKCTAISGDLHLTPVAFKGDSTFRTPDLPRELEITLKYETIGFELLICAMPDNTDL 417
 423 SVFQNLQVIRRIILNAGYSTLTGIGISWLGSLREISGLALLHHNHTLCVHTVPM 482
 418 HAFENLEITRTRTOHGFSLAVGLNTSLGLSLKEISGDVYISGNRLCTANTINM 477
 483 DQLEFNPQALHTANREDECEVGEGLACHOLCARHCWGPPTQVCNCSQFLRQCEVE 542
 478 KKLFTGPQXTRKIMNNAEKCKAVNHYCNPLCSGEGCWGEPRDVCSCQVNSRGRECE 537
 543 ECRVYGLPREYVNAHRLPCHPECQPNQSVTCFGEADQVCANHYKDPFCVAPRS 602
 538 KCMILEGPRFSEVENSECICQHPCLDQANNITGTGQPDNCCICAHYIDSPHCVKTCPA 597
 603 GVKPDLSTYPIWKFPEDEGACQPCPINCSTHSCVDLDGCPAEQASPLTISVAVGIL 662
 598 GIMGENNTL-VKRYADANNVCHLCHANTCYCAGPLOGCEVWSPGRIPIELATGIVGL 656
 663 LVVLGVVFGI-LIKRQOKRKTYTMRRLLOETLEVEPLTSGAMPNOQMRILKEFN 721
 657 LFIIV-VALGIGLEFRRRHIVKRTLRLEORELVEPLTSGAMPNOAHRIKE----- 711
 722 TVSEFLARPKYSASHLETVYKGIWIPDENYKIPAIIVLEBNTPSKANKELDEAYMA 781
 712 --TEPKKRVUSGAGFTVYKGLMPESEKXKIPVATKELDEANSPKANKELDEAYMA 769
 782 GVGSPTVSRLLIGICTSTVOLVTQIMPGCLLDHYENRGLSGQDLLNMCQIAKNSY 841
 770 SVDNPHVCRLLGICLTSTVOLITQIMPGCLLDHYENRGLSGQDLLNMCQIAKNSY 829
 842 LEDVRLVHDLAARVLYKSPNHYKITDPGLARLLIDETRYHAGCGVPIKMMALLESIL 901
 830 LEDRLVHDLAARVLYKSPNHYKITDPGLARLLIDETRYHAGCGVPIKMMALLESIL 889
 902 RRRFTHQSDWVSYGTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYIM 961
 890 HRIYHQSDWVSYGTVWELMTFGSKPYDGIIPASDISSILEKGERLPQPICTIDVYIM 949
 962 VKCMWIDSECRPRELVSFESKMAARDQRTVIG-NEDDLQPASGLDSTFRSLLEDMD 1020
 950 VKCMWIDSECRPRELVSFESKMAARDQRTVIG-NEDDLQPASGLDSTFRSLLEDMD 1009
 1021 GDLVDAEELVPOQGFCEPDPAFGAGVHHRRHSSSTRSGGDLTLLESEEEARSP 1080
 1010 ELDVDAEELVPOQGFCEPDPAFGAGVHHRRHSSSTRSGGDLTLLESEEEARSP 1035
 1081 LAPSEAGSDVFDGLGMAKAGLOSLFTHPSPLQRYSEDPVFLPSET--DGVAAPLT 1138
 1036 LLSLSLSTASN--NSTVACINRNSGCRVKDPAFLQRYSSPFTGAVTNDIDDAFL 1087
 1139 GSPQPEYVNPQVROPSPREPGLPAARPAAGATLERAKTSLSPGKNGVAVKQVFAFGAVE 1198
 1088 --PVEIYVNO-SVPRKPAQSONPYVNHQPLHP-----APGRDLHYCN-PRSNVAV 1154
 1199 NPEYL-TPQGAAPQPHPPAPFAFDNLVYWDQ-----DP-----PERGAPDS 1241
 1135 NPEYLTAG-----PTCLSSGFNSPALMIOGSHQSLMDNDYQODFFPKETKENG 1185
 1242 TEKGTPTAENPEYGLDVP 1260
 1186 IFKG-PTAENAYLKVAPP 1203

RESULT 6

TCHLV

Internal growth factor receptor precursor - chicken
N:Contains: protein-tyrosine kinase (EC 2.7.1.112) etbB

C:Species: Gallus gallus (chicken)

C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C/Accession: A27720 A00643

R/Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart

Mol. Cell. Biol. 8, 1970-1978, 1988

A>Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A:Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720

A:Molecule type: mRNA

A:Residues: 1-1223 <LAX>

A:Cross-references: GB:M20386

R/Nilsen, T.M.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.

Cell 41, 719-726, 1985

A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro

A:Reference number: A00643; MUID:85228222; PMID:2988784

A:Accession: A00643

A:Molecule type: mRNA

A:Residues: 585-1223 <NLS>

A:Cross-references: GB:M10066

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor

C:Specific protein kinase

F:1-30/Domain: signal sequence #status predicted <SIG>

F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>

F:31-654/Domain: extracellular #status predicted <EXT>

F:81-367/Domain: EGF receptor extracellular domain repeat <EE1>

F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>

F:655-677/Domain: transmembrane #status predicted <TM>

F:678-1223/Domain: intracellular #status predicted <INT>

F:719-984/Domain: protein kinase homology <KIN>

F:727-735/Region: protein kinase ATP-binding motif

F:136.702,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #

F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted

F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F:754/Active site: Lys #status predicted

F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.08; Score 3013.5; DB 1; Length 1223;

Best Local Similarity 47.58; Pred. No. 7.2e-117;

Matches 620; Conservative 177; Mismatches 354; Indels 153; Gaps 28;

QY 8 RWGLLALLPPGAA-----STOYCTGDMKRLPASPETHLDMRLHLYGCGVGNIE 61

DB 13 RGAALVLLLVGALCAVEKVCQOTNKKLTQLGVEDHFTSLQRMVNGCVLSNIE 72

QY 62 LLYLPTNASLFLDIOEVGYVLIHNOVROVPLORLIRVGTOLFEVYALAVLDNGD 121

DB 73 ILYVEHNRDLTFLKTIQEVAGYVLIANVDVIPLEMLQIRGVLYDNSFALAVSNYH 132

QY 122 PLNNTPTVGTASPGGRELRLRSITELKGVLIQRMPOLCYODITIMNDIPIKNNOLAL 181

DB 133 -MKTKQ-----GLNELPMKRSELTLNGSVKISNNPKLCNMDYVLMNDIDTSKR-PL 182

QY 182 TLID-TNRSRACHPCSPMCKGSRGWSESSDCSLTRTVAGGCA-RCKGPLPTDCHQ 239

DB 183 TVLDFANLSSCPKCHNCTEDHGWGEGNQOTLTFTVICAQCGSRGKGVSDCHNQ 242

QY 240 CAAGCTGPKHSDCLACIHPHSGICELHCPALVTYNTDTESHPNREGRTFGASCYTAC 239

DB 243 CAAGCTGPRSDCLACKRFRDDATKCTCPLVLYNPTTYQMDVNBGKYSFGATCVRC 302

QY 300 PUYVLTSDVGSCTLVCLHNOEYTAEDGTORCEKSPCARVCGYGLAMQYIKANSKFIGI 359

DB 303 PHNYVVTDHGSCVRSCTDITYEV-EEHGVAKKCKKCGSLGKVCNGIGIKLKLIS-INA 360

QY 360 TELF-PAGCKKIFGSLAFLESPFDGDPASNTAPLQPEQLQVFTLEBITGYLYISKMPDS 418

DB 361 TNIDSPKCTKINGVSVLTPVAFGDATFKTLPLDPKKLDVFRYKELISGLLLQAMPDN 420

QY 419 LPDLSPVONLOVIRRIIHNAGYSITLOGLSWGLRSABLSGLLIHNHTYLCVH 478

DB 421 ATDLAFLNLEIRRTKHOGQYS-LAVNLKIQSGLSLKEISGDIAIMKNKYLQYAD 480

QY 479 TVPMQQLFRNPHQALHTANRPEDECVBGLACHQGLRGKCGWGPFCVVCOSQFLGQ 538

DB 481 TNMMSLRATOSQKTKIIONNRKNDCTDRHVCDELCDVCGKMGCPFHCSGRFFSNK 540

QY 539 ECVEBCRYLQGLPREYVNAHRLCPHCEQPNQNG--SVTCFGEPAQCVAACHYKDPF 595

DB 541 ECVKQCNILQGEPRFERDSKCLPCHSECLVONSTAYNTCSGSPBDHCKAHFIDEPH 600

QY 596 CVARCPGKPDLSMPIMKFPDEGACQPCPINCITHSCVDLDDGCAEGRASPLSTIV 655

DB 601 CVKACPAGVGENDTL-VMKYADANAVQQLCPNCTROCKGDLGCP--NGSKTPSLA 656

QY 656 SAVV-GILLVVLGVVFGILLIKRQOKIRKTYMRLLQETLVEBTLFSGAMPNOQMKRI 714

DB 657 AGVVGGLCLVVGIGIGLYLRR-HVIRKKTLLRLDLRELVEBTLFSGEAPNOAHIRI 715

QY 715 LKEPNNFTVSEFWLRVPKYSASHLETYVYGIWIPDGENYKIPATVYLENNSPKANKEL 774

DB 716 LKE-----TEFKKVKVLSGAFGVYKGLWIPSEKVKIPVAIKELREATSPKANKEIL 769

QY 775 DEAYVAGVGSVYVRLIGICLTSTVQLVTCQMPGCLLDHRENRGLSGODLNMCMQ 834

DB 770 DEAYVMAVDNHNVRRLIGICLTSTVQLTQMPGCLLDYIRKENDIGSYLYLNCVQ 829

QY 835 IAKGNSYLEDVRLVHRDLAARVNVKSNHYKIDTFGLARLLIDETRYHADGKVPYIKW 894

DB 830 IAKGNVYLEBRLLVHRDLAARVNVKTPQHVTITDFGLAKLGADEKRYHAGCKVPYIKW 889

QY 895 MLESILRRPFGHSDVMSYGTWELMTFGAKVPYDGI-PAREIPLLEKRELLPOPPCT 954

DB 890 MLESILRHYTHQSDVMSYGTWELMTFGSKPVDGI-PASEISSVLEKGERLPOPPCT 949

QY 955 IDVWIMVYKCMWIDECPRPRELISESRNARDPQRFVQI-NEDLGPASLSTFFRS 1013

DB 950 IDVWIMVYKCMWIDDSRKPPELLAEFSKVARDPRLVLVIGDERMHLPSPTDSKFFRT 1009

QY 1014 LLEDMDGDVDAEYLYVPOGCFPCPDPAAGCAVHHRHSSSTRSGGDLTGLPESE 1073

DB 1010 LMEEDMDVIDADADYLYVPHOGFF-----NSPST----- 1038

QY 1074 EEARSPPL-----ASEGASGVDFPDGLGMAKGLQSLPTHPSPPLQRYSEDPVPLPS 1128

DB 1039 ---SRPLISSLSLATSNNSATNCTD-----RNGQGHPRFEDSFVQRYSSDPNGNLE 1087

QY 1129 ET-DGYVAFLTCSPOPEYVNOVDYRPOPSRREGPLPAARPAGATLERAKTSLPGKNGV 1186

DB 1088 ESIDDGFL-----PAPEYVNG--LMPKPS-----TAMVQNOI 1118

QY 1187 VKDVF-----AFGAVNPEYVITPQGGAAPOHPHPPASPAFNDNYTYMDQ 1231

DB 1119 YNNISLTAISKLPMDSRVQNSHSTVANDPEYV-----NTNOSPLAKTVFESSPYMIQ 1170

QY 1232 -----DPEP-----RGAPSTFGTPTAENPEYGLDVP 1260

DB 1171 SGNHQINLDNPYQDDPLPNEFKRPGKLKVPAAENPEYLVRAAP 1214

RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C:Species: Homo sapiens (man)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999

C/Accession: A47253

R/Florman, G.D.; Colausco, J.M.; Whitney, G.S.; Green, J.M.; Carleton, G.W.; Foy, L.

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A>Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide

Db 763 VEFMDEALIMASNDHPHLVRLGLGCLSPITQLVTQGMPHGCLLEVEHEKNDIGSQLLIN 8

E:726-1866/RDQuCC:	kinase-related translocating protein	(TU)	#scatrus Predicted
F:707-872/RDQuCC:	protein kinase homology <KIN>		
F:715-723/Region:	protein kinase ATP-binding motif		

Query Match	37.8%;	Score 2592;	DB 1;	Length 1166;
Best Local Similarity	44.1%;	Pred. No. 1.5e-99;		

Matches 563; Conservative 167; Mismatches 396; Indels 150; Gaps 32;

QY 4 AALCRKGLLALLPRAAST-----QYCTDMKRLRPARPETHLMDLRLXGCGVQON 59
 DB 8 AALQ--LLLVLSISCCSTDPDRKVCQGISNQM--LDNHYLKKMKMYGCVNVL 62
 QY 60 LETTLPNTASLFDLQIOVGVYLIANOVROVPLQRLRIVRGTOLEFEDYVALVDN 119
 DB 63 LETTYQENQDLSFQIOVGVYLIANNEVSTIFLVLRILRGNTLEGNFTLLVMN 122
 QY 120 GDEPLNTTGYTASPGGLRELOLRSTELKGVLIQRNPOLCYQDTILMKDIFHKNOL 179
 DB 123 YOK-NPSSP--DYVYGLKQLOLSNLTLEISGVKXSHNPLCNWETIMWIDVKTSP 179
 QY 180 ALTLIDTNRBRACHPSPCKSGRCSWESSEDCQSLRTVYAGGC-ARCKGLPDPCCHE 238
 DB 180 TMLLIPHAFFEOCKDHGCVNCSWAPRGHCQTKLLCAQCNRRRRGKRPIDCCNE 239
 QY 239 QCAAGCTGPRHSDCLALHFNHSGICELHCPALVYNTDTFESMPNDEGRYTPGASCTA 298
 DB 240 HCAGGCTGPRATDCLACRDFNDGTCTKDCPPKIDYSHQVNDPNIKYTFGAACVKE 299
 QY 299 CPVNLSTDVSGCTVLCPLHNOBVTAEDEGTORCEKSKPCARVCYGLQVYKANSKEIG 358
 DB 300 CPENYVTE-GACVRCSCSAGMLEVD-ENGRSKCPDGVCPKVCDEIGISL-SNTIAVN 356
 QY 359 ITEL-EPAGCKKIFGSLAFLPESFDQDPASNTAPLOPQLOVETLEITGYLYISAMPD 417
 DB 357 STVRSFSNCTKINGDILINRNSFEEDPHYKIGTMDPEHMLNLTVEKITGYLVIMWPE 416
 QY 418 SLPDLSPQNLQVIRIRILHNGAYS-LTIQGLISLGLRSLRELQSLALHNTLCF 476
 DB 417 NMTLSVFNLEIRRTTFSRGSFVVVQVRLQVLRSLKEVASGVNLIKNTLQLEAY 476
 QY 477 VHTVPDQFLRNHQALHTANRBEDECEGLAQHQLCARGHCWPCPTQCVCNSQFRL 536
 DB 477 ANTIMRRLPRSDQGISYDART-----ENQTNNESEDCGW-PGRTMVCSLHYDR 528
 QY 537 GQCEVECKVLQGLPREYVNAHRLCPHECCOPQNSVTCFPEADQVCAAHYNDPPC 596
 DB 539 GGRVASCNMLQOEPRBAOVGRCVQCHQECVQDLSLTCYGGPANCKSAHPDQPC 588
 QY 597 VARPSPGVPLDSYPMKFPDEGACQCPINCTHSCVDLDDKGPAPQRASPLTSIYS 656
 DB 589 ITRCHGILGDTL-IMKXADKMGCCQCHQCTGCGGPGISGRCG-IVSHSLAVG 646
 QY 657 AVVGLLVLVGVFGILIKRQOKIRKXTMRLLQETLVEPLTPSGAMPQAOQRIK 716
 DB 647 LVSGLLITVALLIVLLRRRIK-RKRTIRCLQEKELVEPLTPSGAQAQAFRLIK 705
 QY 717 --EENNFTVSPFWRKVASHLETTYKGIPIPDGENYKIPVAKVLENTSPKANKETL 774
 DB 706 ETEFKDRV-----LQSGAFGYKGMNPDGNNIPAIKIKLRATSPKVNQEV 757
 QY 775 DEAYVWAGSPYVSRILGICLTSTVOLTOLMPYCLLDHYRENGRLGSDLLNMCWQ 834
 DB 758 DEAYVWASVDHPHVCRLGICLTSAVOLTOLMPYCLLDYVRQGER-CQGMNLNMCWQ 817
 QY 835 IAKGASYLEDVRLVHNDLAARVLYVSNHVKITDPSGLARLLDIDTEHAQGVPIK 894
 DB 818 IAKGNYLEERHLYHNDLAARVLYLKNPNHVKITDPSGLKLTAEKEVOAGGVPIK 877
 QY 895 MALESILRRFTQSDVWSYGVVWELMTFGAKPYDIPAREIPDLLEGSELPPIC 954
 DB 878 MALESILQMTYTHQSDVWSYGVVWELMTFGSKPYDIPAKELASYLENGERLPIC 937
 QY 955 IDVYIMVCKMIDSCRPRELVEUSERMAADPORFVYIQNEDLGPASPLDSTFYRSL 1014
 DB 938 IEVYIMVCKMIDPSSRPRELVEUSERMAADPORFVYIQNEDLGPASPLDSTFYRSL 994
 QY 1015 LEQDDKGLVDAEYLVPOQGFPCPDPAAGAGMVAHRRSSSTSGGDDLTLGLEPSRE 1074
 DB 995 LSSDD--DYVDADEYLLPYKRI-----NNGGS----- 1019

QY 1075 EAPRSPLASEGASDVVDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPV-PLPSETDY 1133
 DB 1020 -----EPCLPPYGH-----PVRENSTLRNISPDTQALAEKDLGDH 1055
 QY 1134 VAPLTPSDPEPYVNPDPVPP-----PSPRE-----GPL-APAPGATLERAKTLS 1180
 DB 1056 -----EYVNPQGETSSRLSDIYNPNYEDLTDQMGVSLSSQEAETNFSRPEYLN 1105
 QY 1181 PGKGVVVDVAFPGAVENPEYLPPOGAAQHPRPAPSPAFNLVYWDPPERGAAP 1240
 DB 1106 TNQNSL--PLVSSGSMDDPY---QAG-----YQAF-----LPQTGALT 1140
 QY 1241 STFKGTPTAENPEYLG 1256
 DB 1141 GNGMFLPRAENLEYLG 1156

RESULT 9

A36223
 Kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
 C.Species: Homo sapiens (man)
 C.Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
 C.Accession: A36223, 159164
 R.Kraus, M.H.; Issing, W.; Mikl, T.; Popescu, N.C.; Aaronson, S.A.
 Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
 A.Title: Isolation and characterization of ERBB3, a third member of the ERBB/epider
 A.Reference number: A36223; MUID:90083234; PMID:2687875
 A.Accession: A36223
 A>Status: preliminary
 A.Molecule type: mRNA
 A.Residues: 1-1342 <KRA>
 A.Cross-references: GB:M29366
 R.Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
 Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
 A.Title: Molecular cloning and expression of another epidermal growth factor recept
 A.Reference number: 159164; MUID:90311312; PMID:2164210
 A.Accession: 159164
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: mRNA
 A.Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
 A.Cross-references: GB:M34309; NID:g183990; PID:AAA35979.1; PID:g306841
 C.Genetics:
 A.Gene: GDB:ERBB3; HER3
 A.Cross-references: GDB:119880; OMIM:190151
 A.Map position: 12q13-12q13
 C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
 C.Keywords: ATP; phosphotransferase
 F.707-972/Domain: protein kinase homology <KIN>
 F.715-723/Region: protein kinase ATP-binding motif

Query Match 34.1%; Score 2334.5; DB 2; Length 1342;
 Best local similarity 39.6%; Pred. No. 6.4e-89;
 Matches 523; Conservative 195; Mismatches 460; Indels 143; Gaps 36;

QY 10 GLTALLPREGAA--STQVCGTDMKRLRPARPETHLMDLRLXGCGVQONLELTPT 67
 DB 11 GLTSLAAGSEVNSQAVCPETLNGSLVGTGAENQYOTLYLVERCEVWGNLEIVLTGH 70
 QY 68 NASLSPFDLQIOVGVYLIANOVROVPLQRLRIVRGTOLEFEDYVALVDNPGPLNNTT 127
 DB 71 NADLSFLQWIREVGVYLVANNEFSTLPRLRVRVGRQYVDFKFAFVM-----LVNNT 125
 QY 128 PVTGASPGGLRELOLRSTELKGVLIQRNPOLCYQDTILMKDIFHKNOLALTLDTN 187
 DB 126 ---NSHALQLLTLTLEISGVYIEKNDKCHMDTIDMDRIVDRD--AEIVKD 178
 QY 188 RSRACHPSPCKSGSRCSWESSEDCQSLRTVYAGGC-ARCKGLPDPCCHEQCAAGCTG 246
 DB 179 NGRSCPCHVCKG-KCMGPGSEDCQTLTKICAPQCNHGFSGNPNQCCHEGAGCGSG 237
 QY 247 PKASDCLALHFNHSGICELHCPALVYNTDTFESMPNDEGRYTPGASCTACVYNYLST 306
 DB 247 PKASDCLALHFNHSGICELHCPALVYNTDTFESMPNDEGRYTPGASCTACVYNYLST 306

```

Db      238 PDITDPCACHHNDGACVPCRPQPLVYKTLTFLQLEBNPTKTYQGVCAASCPNHFV-V 296
Qy      307 DWGSCCTLVCPHNOEYTAEDGTQCEKSKPCARVGYIGMOYIKANSKF--IGITELE- 363
Db      297 DQNSCVACPPDKMEVD-KNGLKCEPCGGLCPKACEGTG-----SGSRFQYVDSNIDG 350
Qy      364 FACKCTIFGSLAFLEPFDDPASNTPALQPELOQVPELLEITGYLYISAMPDLSLDLS 423
Db      351 FVNCCTKLGNDLITGLNDPMMKIPALDEKLVNFRVREITGTLNQSMPPHANFS 410
Qy      424 VFQNLQYIRGRILHNGAYS-LTLQGLGISWGLRSRELQSGGLALIHNTLCEFTVPM 482
Db      411 VFSNLTITGGRSILYNGFSLIMKLNLTSLGFRSLKEISAGRIYISANRQLCYHSLNM 470
Qy      483 DQLFNPHQALLHTA-NRPEDEVGELAGHQCACGHCHGPERPTQCVNSQPLRQECV 541
Db      471 TKVLRGTEERLITKNRPRRDCVAGKVDPLCSSGCGPFGQCLSRNRSRGVCV 530
Qy      542 EECRVTLQGLPREYVNAHCLPCHPECPQNGSVTCFGEPAQCVACAHYKDPFCVARGP 601
Db      531 THCNFLNGEPRERFAHEAFCSCHPECPMEGTATCNGSGSDTCAQCHPRDGHCVSSCP 590
Qy      602 SGVKPDLSTYMPMKFPEBEGACOPPCINCHSCVDLDDKCPABQRA---SPLTIVSA 657
Db      591 HGVLG--AKGPIKYRPPVQNECRCHENCCTQCKGPELQDLQTLVLTIGKTLTALTY 648
Qy      658 VVGLVYVGVVFGILIKRQOKIR-KYTRRLQTELEVEEPTPSGAMPNOAMRIK 716
Db      649 IAG--LVVIFMGLGTFLYKRGRIQKRAMRYLEGESELEFLDLS-EKANKVLAIRK 705
Qy      717 EFNNFTVSEWLRYVKSASHL-ETVYKGINIPQENYKIPVAIKVLENTSPANKELID 775
Db      706 ETE-----LRKIKVLGSGVFGTVHKGWIPGESIKIPVCKVIEDKSGROSQAVLD 758
Qy      776 EAYMAGVSPYSLRILGISTVQVLTQMLPQGLDHYRENRGLSGODLNNCMOI 835
Db      759 HMLAIGSLDHAHITRLGLCPGSSQLVTOYLPGLSLDHYRQKRLGQLLNNGVQI 818
Qy      836 AKGMSYLEDEVLYHRLAARNVLYKSPNHVKITDFGLARLLDIDETRYHAGGKVDIKM 895
Db      819 AKGMYLLEHGMVHRNLAARNVLLKSPQVADDFGVADLPPDDKQLTSEAKTEIKM 878
Qy      896 ALBSILRRFTHSDVMSGVTVWELMTFGAKPYDGIPIREIDLEKGRLOPPICIT 955
Db      879 ALBSIFRGKITHSDVMSGVTVWELMTFGAEPYAGRLAEVPLLEKGERLQPOICTI 938
Qy      956 DVYVIMVCKMWIDSECRPRFRELVSFSRMARDPQREYVLIQNEDLGPA--SPUDSTFYR 1012
Db      939 DVYVIMVCKMWIDENIRPTEKELANFTRMARDPRYLVIKRES-GRGIAPGPEHGLTN 997
Qy      1013 SLEDDDMGDLVNAEELVPOQGFCDPAPAGAGMWHHRSSSTSSGGDLITLGLER- 1071
Db      998 KKELEVELEPDLIDLLEAED-----NLATTLTASALSLPVGTL 1038
Qy      1072 SEEEAPRSPPLAPSEAGSDVFDGLGMAKGLQSLPTHD-PSPLQRYSEDPTVPLP- 1127
Db      1039 NRPRGSGSLSPSSGCV-MPMNGKGLSGCOESANSSGSERCPRVSLH-----PMRGC 1091
Qy      1128 ---SETDGYVA-----PLTCSPOPE---YVNOPDVROPSPRRSGP- 1162
Db      1092 LASESSEGHVTSAEALQEKVSMCRSRSPRFRGDSAYHSORHSLTLPVPLSPFGL 1151
Qy      1163 ---LPAARPAATLEERAKTSLP-GKNGV-----KQVFARGAIVENPEYLTQ 1206
Db      1152 EEDLVNGVWPDTHLKKTTPSSREGLTSSVGLSVLGEDEED-----EVEYINNR 1203
Qy      1207 GGAAPQHPHPAPSPAFDNLVYND-----QDPERGAPSTFTFGTPTAENPEY 1254
Db      1204 RHRSP-FHPRPSSLBEIAGEYMDVGSDLASLQSTQSCPLHVPVPIPTAGTTDDEYEX 1262
Qy      1255 L 1255
Db      1263 M 1263

```

```

RESULT 10
JC4387
Epidermal growth factor receptor homolog precursor - rat
N.Alienate names: ErbB3 protein; HER3 protein
C.Species: Rattus norvegicus (Norway rat)
C.Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C.Accession: J04387
R.Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A.Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant prot
A.Reference number: J04387; MUID:96096535; PMID:8522190
A.Accession: J04387
A.Molecule type: mRNA
A.Residues: 1-1339 <HEL>
A.Cross-references: GB:029339; NID:9915389; PID:9915390
A.Experimental source: Liver
A.Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res
C.Comment: This protein is a functional heregulin receptor that transduces signals
C.Genetics:
A.Gene: ErbB3
C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C.Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-976/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Ty
Query Match 32.8%; Score 2248.5; DB 2; Length 1339;
Best Local Similarity 39.8%; Pred. No. 2,2e-85;
Matches 514; Conservative 171; Mismatches 440; Indels 167; Gaps 37;
Qy      3 LAALRWGILLALLPPGAA---STQVCTGTDMKRLRPSPEHIDMLHLYOGCQVQCN 59
Db      7 LQVLC---FLTLARGSEMNSQAVCGTNGSLVTDADNQYOTLYKLEKCEVYNGN 62
Qy      60 LETLTLPTNASLSPFQDIEQVGVLYLHANGVQVPLQRLRIVRGSTOLFEDNYVALAVLDN 119
Db      63 LEIVLTGNHADSFLQWIREVTAYLVAMNESVPLNLRNVRGTQYVDGKFAIFVM-- 120
Qy      120 GDPPLNNTPTVTGASGGLRELOLSLTLEILKGVLTQKRNPOLCYDOTTLMKDIPKRNOL 179
Db      121 ---LNYNT---NSSHALROLKFTQLTLEILSGVYIEKNDKICMDDTIDMDIVVR-- 170
Qy      180 ALTLIDTKRSRACHPCSPMKSGRCMGSESSDCOSLRTVCAGG-ARCKGPLPDDCCH 238
Db      171 GAEIVYKXNGANCPCHEVCKG-RCMGEGPDDCOLITTTICAPQCNAGCFGPNPOCHD 229
Qy      239 QCAAGCTGPKXSDCLALHPNHSIGICELHCPALVYNTDTESMNPREGRYTFGASCYTA 298
Db      230 ECAGGCGSGQDDTDCACRRFNDGACVPCRPQPLVYKTLTFLQLEBNPTKTYQGVCAAS 289
Qy      299 CPYVNLSTDVSGCTVCPHNOEYTAEDGTQCEKSKPCARVGYG--GMQYIYANSKF 356
Db      290 CPNHFV-VQDTFCVACPPDKMEVD-KNGLKCEPCGGLCPKACEGTGSGSYQVDSN 347
Qy      357 IGITELPAGCKKIFGSLAFLEPFDDPASNTPALQPELOQVPELLEITGYLYISAMP 416
Db      348 ID----GFVNCCTKILGNDLITGLNVDPMMKIPALDEKLVNFRVREITGTLNQSMP 403
Qy      417 DSLPDLSTYQNLQVIRGRILHNGAYS-LTLQGLGISWGLRSRELQSGGLALIHNTLHC 475
Db      404 PMHNHFSVPSNLTITGGRSLYNGFSLIMKLNLTSLGFRSLKEISAGRIYISANRQLC 463
Qy      476 FVHTVPWQLFRNPHQALLHTA-NRPEDEVGELAGHQCACGHCHGPERPTQCVNSQ 534
Db      464 YHSLNMTNRLRNGPEERLDIKYRPLDECLAGVCPPLCSSGCGMGPARGQCLSGN 523
Qy      535 LFGQCEVECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEPAQCVACAHYKDP 594

```

Db 524 SREGVYTHCNFLGSEPREFEVHNAQCFSCHEPCLPMEGTSTVNGSGSDACARCAHPFDDG 583
 QY 595 FCYARCPSSGVKPLDSTVPIKPFPEBACQPCPINCNSHC--VDLDKXGAPAGRSAPLT 652
 Db 584 HCVNACBPHILG--AKSPIKYKPDQNECPCHENCIOGNGPLOCGLGQAVLMSKPH 641
 QY 653 SIYSAVVGILLVVLGVVFGILIKRQOKIR-KYIMRRLQETELVEPLTPSGAMENQAO 711
 Db 642 LVIAVYTG--LAVILMLTGSFLVWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVL 698
 QY 712 MRLIKERNFTVSFWLAVPKYSASHL-ETVYKGIWIDGENVKIPVAKYLRNTPSKAN 770
 Db 699 ARIFKETE-----LRKLYLGSVGVYKGIWIDGESIKIPVCIKYEKSGQSF 751
 QY 771 KEILDEAVYVMAVGSFVYSRLIGLCTSTVOLVTLQMPYGCILDHVENRGLSGODLNL 830
 Db 752 QAVTDHMLAVOSLDHAIIVLLGICPSSSLQVLYQYPLDGLDHVQNHETIGPOLLN 811
 QY 831 WCMQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKTITDGLARLLDIDETVHADGKV 890
 Db 812 WGVQIAKGMVYLEHSHVHRDLARNVLMKSPSCQVADFGVADLPPDDKOLLHSEAKT 871
 QY 891 PIKMMALSTLRREFTHOSQVWSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPOR 950
 Db 872 PIKMMALSTLHFGKITHOSQVWSYGVTVWELMTGAKPYDGIPIAREIPDLEKGERLPOR 931
 QY 951 PICTIDVYMIWVKCMIMIDSECRPRFRELVSFSSMAADPORFVYIQNEDLGAPSPLDSTF 1010
 Db 932 QICTIDVYMIWVKCMIMIDSECRPRFRELVSFSSMAADPORFVYIQNEDLGAPSPLDSTF 988
 QY 1011 YRSLIEDDDMDLDVDAEYVLPVQGFPCPDPAAGMWHHRHSSSTRSGGDLTLGLE 1070
 Db 989 EPSVLTTEL-----QSAELEPEL-----DLDDLE 1014
 QY 1071 PSEB-----EADPSPLAPSEG-----AGSDVFDGDLGMAAK 1102
 Db 1015 AEEBGLATSGSALSLPTGLTLTRRGSGSLSPSSGVMNQSLGEACLDASVLTGREQ 1074
 QY 1103 GLGSLPHYDPSPLORYSEDPVLPISLTDGYV---APL-----TC-----SPOPE-- 1144
 Db 1075 FSRPISLH-PIPRGR-----PASESEGHVTSSEAELOEKVSVCSRSRSPRPRGCD 1126
 QY 1145 --YVNOPDVRPQPSPREGR-----LPAPAPGATLESAKTLSP-GKKGVV-- 1187
 Db 1127 SAYHSQRSLTLPVTPSPGLEEDGNGYVMPDTHLRGASSSREGTLSSVGLTE 1186
 QY 1188 --XDVFAFGAVENPEYLTPOGGAAPQHP 1216
 Db 1187 EDED-----EEYEMNRKRKRGSP-PRPP 1209

RESULT 11
 TVFVTV
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
 N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
 C:Species: avian leukosis virus, AIV
 C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
 C:Accession: B00643; A00643
 R:Nilsen, I.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
 Cell 41, 719-726, 1985
 A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:85228222; PMID:2988784
 A:Molecule type: mRNA
 A:Residues: 1-658 <NLI>
 A:Cross-references: GB:M0066; GB:M1981; NID:g211749; PIDN:AAA4763.1; PID:g211750
 A>Note: in Genbank entry CHKRBEP, release 109.0, the source is designated as Gallus gal
 C:Comment: This protein is synthesized as a gag-env-erbB protein.
 C:Genetics:
 A:Gene: gag-env-erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 P1-6/Product: gag protein (fragment) #status predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>
 F:60-698/Product: protein-tyrosine kinase erB #status predicted <ERB>
 F:194-459/Domain: protein kinase homology <KID>
 F:202-210/Region: protein kinase ATP-binding motif
 F:229/Active site: Lys #status predicted

Query Match 24.6%; Score 1685.5; DB 1; Length 698;
 Best Local Similarity 50.1%; Pred. No. 1.6e-62;
 Matches 362; Conservative 84; Mismatches 145; Indels 131; Gaps 19;

QY 578 GPEADQVCANHYKDPPEVCARCPGSKVPLDSTVPIKPFPEBACQPCPINCNSHCVDL 637
 Db 60 GP--DHCKCAHFIIDGPHCVKACPAGVAGENTL-VKADANACQLCHPCTROCKRP 116
 QY 638 DDKGPAEORASPLTSYSAVY-GILLVVLGVVFGILIKRQOKIRKYIMRRLQETEL 696
 Db 117 GLEGCC--NGSKTBSIAAGVVGILLCVVGLIGLYLRR-HIVRKTLRLLOEREL 172
 QY 697 VERPLTPSGAMPQAKRILIKERNFTVSFWLAVPKYSASHLETVYKGIWIDGENVKIPV 756
 Db 173 VERPLTPSGAPQAHRLIKR-----TEFKVXVLSGAFGVYKGLMIPGEEKVXIV 226
 QY 757 AIKYLRENTSPANKELIDEAVYVMAVGSFVYSRLIGLCTSTVOLVTLQMPYGCILDHY 816
 Db 227 AIKYLRENTSPANKELIDEAVYVMAVGSFVYSRLIGLCTSTVOLVTLQMPYGCILDHY 286
 QY 817 RENGRLSGQDLINCMQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKTITDGLARLL 876
 Db 287 REHKDNGISQYLLNMCVQIAKGMVYLEHSHVHRDLAARNVLYKTPQHAKITDFGAKLL 346
 QY 877 DIDETVHADGKVPIKMMALSTLRREFTHOSQVWSYGVTVWELMTFGAKPYDGIPIARE 936
 Db 347 GADKEKHADGKVPIKMMALSTLRITTHOSQVWSYGVTVWELMTFGSKPYDGIPIARE 406
 QY 937 IPDLLEKGERLPORPCTIDVYMIWVKCMIMIDSECRPRFRELVSFSSMAADPORFVYIQ 996
 Db 407 ISSVLEKGERLPORPCTIDVYMIWVKCMIMIDSECRPRFRELVSFSSMAADPORFVYIQ 466
 QY 997 -NEDLGAPSPLDSTFRLSLIEDDDMDLDVDAEYVLPVQGFPCPDPAAGMWHHRHRS 1055
 Db 467 GDRRMHLPSFTSKFRTLMEEDMDLDVDAEYVLPVQGF-----NS 510
 QY 1056 SSTRSGGDLTLGLESEEAAPSP-----APSEAGSDVFDGDLGMAKGLQSLPTH 1110
 Db 511 PSL-----SRPLLSLSAHSNNNSAINCID-----RNGQGHPR 544
 QY 1111 DPSPLORYSEDPVLPISLTDGYV---APL-----TC-----SPOPE-- 1187
 Db 545 EDSFVORYSSDPTGNFLEESIDGFL-----FAPEYVQ--LMPKKPS----- 585
 QY 1169 AGATLERAKTLSPGKXGVYKDV-----AFGAVENPEYLTPOGGAAPQ 1213
 Db 586 -----TANVQOITVNNISLTAKLPMDSRVONSHSTVVDPEYI-----NT 627
 QY 1214 HPPAPSPAFDNLVYWDQ-----DPPE-----RGAPSTFGKTPAENPEYLGID 1258
 Db 628 NQSPPLAKTVSESSPYIQSGNHQINDNPYQDDPLPNETKPNGLKVPAAENPEYLRVA 687
 QY 1259 VP 1260
 Db 688 AP 689

RESULT 12
 TVYUHV
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain
 C:Species: avian erythroblastosis virus
 C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
 C:Accession: A00644; A38022
 R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
 Cell 35, 71-78, 1983
 A>Title: The erbB gene of avian erythroblastosis virus is a member of the src gene
 A:Reference number: A00644; MUID:84026539; PMID:6133229

A:Accession: A00644
 A:Molecule type: DNA
 A:Residues: 1-604 <YAM>
 A:Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
 A:Rebuture: B. Henry, C. Benalissa, M. Bisette, G. Claverie, J.M. Saule, S. Martin, Science 224, 1456-1459, 1994
 A:Title: Sequencing the erbB gene of avian erythroblastosis virus reveals a new type of A:Reference number: A38022; MUID:64223957; PMID:6328658
 A:Accession: A38022
 A:Molecule type: DNA
 A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>
 A:Cross-references: GB:K02006
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 F:130-395/Domain: protein kinase homology <KIN>
 F:138-146/Region: protein kinase ATP-binding motif
 F:165/Active site: Lys #status predicted

Query Match 23.7%; Score 1622; DB 1; Length 604;
 Best Local Similarity 50.0%; Pred. No. 5.8e-60;
 Matches 348; Conservative 80; Mismatches 136; Indels 132; Gaps 17;
 QY 587 CAHYKDPFCVACPSGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDLDDKGGPAGQ 646
 DB 3 CAHFIDGPHCVKACPAVLENDTL-VRKYADANAQCQLCHPCTRCCKGPGEGCP--- 58
 QY 647 RASPLTIVASV-GLLVVVLGVFGILIKRQOKIRKTYMRLLQETLVEPLTPSGA 705
 DB 59 NGKTFPSIAGVGGGLLVVGLGILYLR-RIYKRTLRLLQERLEVEPLTPSGE 117
 QY 706 MPNOQMRILKEFNFTVSFMLRPVYSASHLETYVKGIMIPGENVKIIVALKVRENT 765
 DB 118 APNOAHRLIKE-----TEFKYKVLGSGAFGTITKGLMIPSEKXIKIVALKELREAT 121
 QY 766 SPKANKILDEAYMAGVSPYVSRLLGICLTSTVQVLTQLMRYGCLLDHRENGRLGS 825
 DB 172 SPKANKILDEAYMASVNDPNHVCRLGICLTSTVQVLTQLMRYGCLLDHRENGRLGS 221
 QY 826 QDLNMCQIAKGSYLEDVRLVHDLAARNVLYKSPNHVKITDFFGLARLLDDETEYHA 885
 DB 232 QYLLNMCVOIAKGNVLEERLVHDLAARNVLYKTPQHKITDFFGLAKLGDDETEYHA 291
 QY 886 DGGKVPKIMMALESILRRFTHOSDVSYGVYVWELMTFSKAPYDGIIPAREIDPLEKGE 945
 DB 292 EGGKVPKIMMALESILRRFTHOSDVSYGVYVWELMTFSKAPYDGIIPAREIDPLEKGE 351
 QY 946 RLPOPICTIDVYIMVKCMWIDSECRPRELVSEFSKARDPQRFVYIO-NEDLGPAS 1004
 DB 352 RLPOPICTIDVYIMVKCMWIDSECRPRELVSEFSKARDPQRFVYIO-NEDLGPAS 411
 QY 1005 PLDSTFYRSLLEDVDDMDLVDAEYVLPQGFPCPDPAAGGVHHRHSSSTRSGGD 1064
 DB 412 PTDSKFYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST----- 449
 QY 1065 LTUGLESEBEARSP-APSEAGSDVFPDGLGMAKGLQSLPHHDSPLQRYIS 1119
 DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHPRVDESFPQRYIS 489
 QY 1120 EDTPVPLPSET--DGYVAPLTCSPQPEYVNOQDVPRQPSPREGGLPAARPAAGTLERAK 1177
 DB 490 SDPTGNFLESIDGGL-PAPEYVNO--LMPKKPSTAM----- 524
 QY 1178 TLSPGKNGVYKDVAF-----GAVENPERILTQGGAAPOPHPPA 1218
 DB 525 -----VQNOINFLSLTAISKLPMDSRYQNSHSTAVIDNPEYL-----NTQOSPL 568
 QY 1219 FSPAFLNLYWDDPPEPGAPPTFGKPTAENPEY 1254
 DB 569 AKTVFESSPYWIGSGNQ-----INLDNDY 594

RESULT 13
 S35745
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
 C:Species: avian erythroblastosis virus
 C>Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C:Accession: S35745
 R:Vennstrom, B
 submitted to the EMBL Data Library, March 1993
 A:Reference number: S35743
 A:Accession: S35745
 A:Molecule type: DNA
 A:Residues: 1-544 <VEN>
 A:Cross-references: EMBL:X12707
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-speci
 F:135-400/Domain: protein kinase homology <KIN>
 F:143-151/Region: protein kinase ATP-binding motif
 F:170/Active site: Lys #status predicted

Query Match 22.9%; Score 1569; DB 2; Length 544;
 Best Local Similarity 52.7%; Pred. No. 7.9e-58;
 Matches 334; Conservative 73; Mismatches 129; Indels 98; Gaps 16;
 QY 578 GPEADOCVACAHYKDPFCVACPSGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDL 637
 DB 1 GP--DHCKCAHFIDGPHCVKACPAVLENDTL-VNRYADANAQCQLCHPCTRCCKGPG 57
 QY 638 DDKCPAEORASPLTIVASV-GLLVVVLGVFGILIKRQOKIRKTYMRLLQETL 696
 DB 58 GLEGGP---NGKTFPSIAGVGGGLLVVGLGILYLR-RIYKRTLRLLQERLEVEPLTPSGE 113
 QY 697 VEPLTPSGAMPNOQMRILKEFNFTVSFMLRPVYSASHLETYVKGIMIPGENVKIIV 756
 DB 114 VEPLTPSGAMPNOAHRLIKE-----TEFKYKVLGSGAFGTITKGLMIPSEKXIKIV 167
 QY 757 AIKVLRENTSPKANKILDEAYMAGVSPYVSRLLGICLTSTVQVLTQLMRYGCLLDHV 816
 DB 168 AIKVLRENTSPKANKILDEAYMASVNDPNHVCRLGICLTSTVQVLTQLMRYGCLLDHV 227
 QY 817 RENGRLGQDLNMCQIAKGSYLEDVRLVHDLAARNVLYKSPNHVKITDFFGLARLL 876
 DB 228 RENGRLGQDLNMCQIAKGNVLEERLVHDLAARNVLYKTPQHKITDFFGLAKLGDDETEYHA 291
 QY 877 DIDETRYADGKVPKIMMALESILRRFTHOSDVSYGVYVWELMTFSKAPYDGIIPARE 936
 DB 298 GADKEXYNAEGKVPKIMMALESILRRFTHOSDVSYGVYVWELMTFSKAPYDGIIPARE 347
 QY 937 ITDLEKGERLPOPICTIDVYIMVKCMWIDSECRPRELVSEFSKARDPQRFVYIO 996
 DB 348 ISSVLEKGERLPOPICTIDVYIMVKCMWIDSECRPRELVSEFSKARDPQRFVYIO 407
 QY 997 -NEDLGPASPLDSTFYRSLLEDVDDMDLVDAEYVLPQGFPCPDPAAGGVHHRHSS 1055
 DB 408 GDERHMLBSPSTKSKYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST----- 451
 QY 1056 SSTSGGGGLVTLGLEPSEBEARSP-APSEAGSDVFPDGLGMAKGLQSLPHHDSPL 1110
 DB 452 PST-----SRTPLLSLSATSNNSATNCIDRNG-----H 481
 QY 1111 DSPPLQRYSEDPVPLPSETGQVYVAPLTCSPQPEYVNOQDVPRQPSPREGGLPAARPA 1170
 DB 482 -----PYREDGFL-----PAPEYVNO--LMPKKPSTAMONQIYNTIS 517
 QY 1171 ATLERAKTLSPGKNGVYKDVAFAGAVENPEYL 1203
 DB 518 LTAISKPLPDRSQN-----SHSTAVIDNPEYL 544

RESULT 14
 S00727
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi

C/Species: avian erythroblastosis virus
 C/Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C/Accession: S00727
 R/Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A/Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
 A/Reference number: S00727; MUID:88217326; PMID:2897102
 A/Accession: S00727
 A/Molecule type: DNA
 A/Residues: 1-545 <SCO>
 A/Cross-references: EMBL:X06943
 C/Genetics:
 A/Gene: erbB
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; phosphotransferase
 F/135-400/Domain: protein kinase homology <Kin>
 F/143-151/Region: protein kinase ATP-binding motif

Query Match 22.8%; Score 1562; DB 2; Length 545;
 Best Local Similarity 52.7%; Pred. No. 1.5e-57;
 Matches 314; Conservative 72; Mismatches 130; Indels 98; Gaps 16;

QY 578 GPEADQCVACAHYKPPFCVAPCPGKPDLSYMDIMKPPDEGACQCPINCTHSCVDL 637
 DB 1 G--DHCMKAFIIGPHVCACPAGVGENDTL-VKXADANAVCQLCHPCTRGCKGP 57
 QY 638 DDKGPAPGASPLTISAVV-GILLVVLGVFGILIKRQKIRKYMRLLOETEL 696
 DB 58 GLEGCP---NGSKTSLAAGVGGILLVVGGLGILYLR-R-HIVRKRLRLLOEREL 113
 QY 697 VEPLTPSGAMPQAOARILKEFNNTVSMLEVPKVASHLETYYKGIWIDGEMVKIPV 756
 DB 114 VEPLTPSGAMPQAOARILKE-----TEFKVKVLGFGAGFTVYKGIWIDGEMVKIPV 167
 QY 757 AIKVLRENTSPKANKELIDEAYVMAGVSPYVRLIGICTISTVOLVQLMPPYGCILDHV 816
 DB 168 AIKELRENTSPKANKELIDEAYVMAGVSPYVRLIGICTISTVOLVQLMPPYGCILDHV 227
 QY 817 RENGRLSQDLINCMQIAKMSYLEDRVLVHRLAARNTLVKSPNHVKTITDGLARLL 876
 DB 228 RENGRLSQDLINCMQIAKMSYLEDRVLVHRLAARNTLVKSPNHVKTITDGLARLL 287
 QY 877 DIDETEYADGSKVYIKMMALLESILRRFTHQSDVMSYGVWELMTGAKPYGIPARE 936
 DB 288 GADEKEYTAEGSKVYIKMMALLESILRRFTHQSDVMSYGVWELMTGAKPYGIPARE 347
 QY 937 IPDLLEKGERLPQPICTIDVYIMVKCMWIDSECRPPREFELVSFSMAADPQRFVVIQ 996
 DB 348 ISSVLEKGERLPQPICTIDVYIMVKCMWIDSECRPPREFELVSFSMAADPQRFVVIQ 407
 QY 997 -NEDLGPAFLDSTFYRSLLEDDDMGDLVDAEYLVPOQGFCCPPAPAGAMVHHIRRS 1055
 DB 408 GDERHMLSPDTSKRYRTLMEEDEMDVDADEYLVPOQGF-----NS 451
 QY 1056 SSTRSGGDDLTLGLEPSESEAPRSL-----APSEAGSDVFDGDLGMAKGLQSLPTH 1110
 DB 452 PST-----SRPLLSLSATSNNSATNCIDRNG-----H 481
 QY 1111 DPSPLORYSEDTVLPSEITGYVAPLTCSPQPEYVNPQVDRPQSPSRREGPLPAAPAG 1170
 DB 482 -----PVREDFL-----PAPEYVNG-LMPKKPSTAMVQNOIYVVIS 517
 QY 1171 AT-TERAKTLSPGKNGVVDVAFAGAVENEYL 1203
 DB 518 LTAISKLPMDSRYN-----SHSTAVDNPEYL 544

RESULT 15
 GQFFE
 N/Contains: growth factor receptor - fruit fly (Drosophila melanogaster)
 C/Species: Drosophila melanogaster
 C/Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999

C/Accession: A00640; A38021
 R/Lytne, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.
 Cell 40, 559-607, 1985
 A/Title: The Drosophila EGF receptor gene homolog: conservation of both hormone bi
 A/Reference number: A00640; MUID:85124611; PMID:2982499
 A/Accession: A00640
 A/Molecule type: DNA
 A/Residues: 1-1330 <LIV>
 A/Cross-references: EMBL:X03054
 R/Madsen, S.C.; Vincent III, W.S.; Bloddeau-Wentworth, D.
 Nature 314, 178-180, 1985
 A/Title: A Drosophila genomic sequence with homology to human epidermal growth fac
 A/Reference number: A38021; MUID:85137938; PMID:2983222
 A/Accession: A38021
 A/Molecule type: DNA
 A/Residues: 1-832-866, 'V', 868-943, 'QTEPLVY' <MAD>
 A/Cross-references: EMBL:X02293; NID:97922; PID:CA26157.1; PID:97929565
 C/Comment: This sequence is tentative because the introns have not been identified
 C/Genetics:
 A/Gene: FlyBase:Egfr
 A/Cross-references: FlyBase:FBgn0003731
 A/Map position: 2 57F
 C/Superfamily: epidermal growth factor receptor; protein kinase homology
 C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; p
 F/1-732/Domain: extracellular #status predicted <EXT>
 F/733-764/Domain: transmembrane #status predicted <TM>
 F/765-1330/Domain: intracellular #status predicted <INT>
 F/808-1072/Domain: protein kinase homology <Kin>
 F/816-824/Region: protein kinase ATP-binding motif
 F/122,300,324,363,518,668,695,700/Binding site: carbohydrate (Asn) (covalent) #sta
 F/774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predi
 F/843/Active site: lys #status predicted
 F/1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status p

Query Match 22.7%; Score 1557.5; DB 1; Length 1330;
 Best Local Similarity 29.0%; Pred. No. 5.6e-57;
 Matches 404; Conservative 178; Mismatches 419; Indels 391; Gaps 42;

QY 80 VQGVYLAHQVROVPLQRLIRVGTOLF-----EDNVYALAVLDNGDPLNNTTPVGTASP 134
 DB 38 ITNYIVIGLDLIPCTLSYRLQIIRGRILFSLVSEKXALFV-----TY 81
 QY 135 GGLRELDRLTELKSGVLIQRNPOLCYDITLMKDIHKNNQALITLIDNRSRACHP 194
 DB 82 SKMTLEIPDLRVLNQGVEFNNYMLCHRRITOWSEIVSNGTDAYNYDFTAPRECPK 141
 QY 195 CSPMKSGSRGCGSSSEDCQSLRTVACAGCA--RCKGPLPTDCHEGCAAGCTGPKXSDC 252
 DB 142 CHECHTG-CWEGEPKRCQCFSLKTCSPQACAGRCYGFKRECHLFCAGCGTGFQKDC 200
 QY 253 LACLHNHSGICELHCPALVYNTDTFESMPNPEGRTTFGASCVTACPVNYLSTDVSGCT 312
 DB 201 IACKNPFDEAVSKEECPKMKYNPPTVLETNPEGKAYAGATCVKECP-GHLLRDNGACV 259
 QY 313 LVCPRLNQEYTAADGTORCGKSKPCARVYGVGMQYIKANSKFIGITEL-----EFAG 366
 DB 260 RSCPDQKMDKGE-----CYPCNGPCPKTC-----PGTVLAHAGNDSFRN 300
 QY 367 CKKIFGSLAFSPESFDG--DPAANTA-----PLQPEQLQVFELEITGTYLISAMPDS 418
 DB 301 CTVIDGNIRILDTQFGSQGVVANYMGPRYIPLDERRRVFTVWEITTYLIEGHHQ 360
 QY 419 LPPLSVFQNLQVIRGRLHNKAY-SLTQGLGISMGLSRLBSGLALIHNTLCPV 477
 DB 361 FRLSLYFRNLTEHGRQLWESMAALAIYKSLYSLEMRNLKQISGSSVVIQNRDLQVY 420
 QY 478 HTVPMDQLFRNPQALHTANRDEEC----- 504
 DB 421 SNIRWPAIQKEPKQKVVAVENLRADLCGRFTLLISVQNNIWHIPIAKREKNNHLSGV 480
 QY 505 ----- 504
 DB 481 QRGRLGSMHGSVPLQELQFQWHLHRLMLYIQVINSITQDKSNEHQLTDACVSPSVPT 540

QY 505 -----VG 506
Db 541 SLTIERARYAIOAGLAMELEQITARSASMRHKTLPAREGQVPRWFLGVACASARAGIA 600
QY 507 EGIA-----CHOLCARGCWGPPTQCVNCSOPLRGCEVEECRVLOGLPREYV---N 556
Db 601 EPLAGAVCRKCHPLCELTCTNYGHEQVCSKCTHYKRRQCETEC-----PADHYTDEE 654
QY 557 ARHCLPCEPCOPONGSVTCFGEBADQVACAHYK-----DPPF-----CVARCPG 603
Db 655 ORECEFORHPEC---NG---CTGPGADDCCKSCRNFKLFDANETGPYVNSTMFNCTSKCPLE 708
QY 604 VK-PDLSTYMPWKFPPEEGACQCPINCHSCVDLDDKCPAEGRASPSTISVASAVGIL 662
Db 709 MHRVNYQYTAIGPY-----CAASPRSSKITAND-----VNIIFITGAV 749
QY 663 LVVVLGVVFGI-LIKRROQKIRRYT--MRLLQETELVEPLTSGAMPNOQMRLIKEFN 719
Db 750 LVPTICILCVTYICROKOKAKKETVQMTALSGRESEPLRPSNIGANLCKLRIYKDAE 809
QY 720 NTFVSFWLAVPKV-SASHLETYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDEAY 778
Db 810 -----LRKGGVLGAGARVYKGVWPEGENVKIPVAIKELKSTGAESESEFLREAY 862
QY 779 VMAGVSSPYVSRLLGICLTSTVOLVTOIMPYGCLLDHYRENGRLGSDLLNMCMQIAKG 838
Db 863 IMASEEHVNLKLLAVCMSSQWMLITQMLPLGCLLDYVRNNRDKIGSKALNWSQIAKG 922
QY 839 NSYLEDEVRLVHRDLAARNLVN---SPNHVKITDFGLARLLDIDETEHADGKVPYKMM 895
Db 923 NSYLEBEKRLVHRDLAARNLVRLLAGEDH---DFGLAKLLSSDSNEYKPAAGKMPIKML 978
QY 896 ALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTI 955
Db 979 ALECIERNRVFTSKSDVMAFGVIMWELTFCGRPHENIPAKDIPDLIEVGLKLEQPEICSL 1038
QY 956 DYYMIMVKCMIDSECRPRFRELVESEFSMARDPQRFVVIQNEIDLG--PASPLDSTFYRS 1013
Db 1039 DIYCTLLSCMHLDAAMRPTFKQLTVFAFAPARDPGRYIALIGDKFTRLPA-----YTS 1091
QY 1014 LLEDD---DMGDLVDAEYLVPOQGFPCDPAPAGAGMWHRRSSSTRSGGDLTLGLE 1070
Db 1092 QDEKDLIRKLAFTTDSSEALAKPDDYLOPKALGFS-----HRTDCT----- 1133
QY 1071 PSEEEAP-----RSP LAPSEAGSDVFDG---DLGMAKAGLQSLPTHDPSPLOYSED 1121
Db 1134 ---DEMPKLNRYCKDPSNNSSTGDDERDSSAREVGGNLR----- 1171
QY 1122 PTYPLPSEEDGVAPLTGSPQPEYVQPDVRPOPSPREGPLPAARPAATLERAKITLSP 1161
Db 1172 --LDLPVDEDDYLMPTCQPGPRNNNNNA-----NPNONMAAVGVAAGYV----- 1214
QY 1182 GNGGVVKDVAFAGGAVENPEYL---TPGGAAPOPH-----PPPAF 1219
Db 1215 -----DLIGVAVSVDNPEYLLNAQTLGVGESPIPTQITGIPVWGPGTMEVKVPMFGS 1267
QY 1220 SP-APDNLYYMD 1230
Db 1268 EPTSSDHEYND 1279

Search completed: July 22, 2003, 09:27:59
Job time : 31.3921 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.6928 Seconds

(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-695-709-14
Perfect score: 6853
Sequence: 1 MEALALCRWGLLALLPFGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6641	96.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5850	85.4	1257	1 ERB2_RAT	P06494 rattus norv
3	5834.5	85.1	1254	1 ERB2_MESAU	O60553 mesocricetu
4	3063	44.7	1210	1 EGFR_HUMAN	P00633 homo sapien
5	3037	44.3	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	2884.5	42.1	1308	1 ERB4_HUMAN	Q15303 mus musculu
7	2865	41.8	1308	1 ERB4_RAT	Q62956 rattus norv
8	2608.5	38.1	1167	1 XMRK_XIPMA	P13388 xiphophoru
9	2343.5	34.2	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2271.5	33.1	1339	1 ERB3_RAT	O62799 rattus norv
11	1879	27.4	1426	1 EGFR_DROME	P04412 drosophila
12	1668.5	24.3	634	1 ERBB_AIV	P00534 avian leuko
13	1622	23.7	604	1 ERBB_AVIER	P00635 avian eryth
14	1595	22.3	703	1 EGFR_CHICK	P13387 gallus gall
15	1552	22.6	540	1 ERBB_AVIEU	P11373 avian eryth
16	1241	18.1	1323	1 LT23_CAEEL	P24448 caenorhabdi
17	1142.5	16.7	245	1 ERB2_MOUSE	P70424 mus musculu
18	700	10.2	1363	1 ILPR_BRALA	O02466 brachyosco
19	670	9.8	1363	1 INSR_RAT	P15127 rattus norv
20	669.5	9.8	1372	1 INSR_MOUSE	P15208 mus musculu
21	665	9.7	1300	1 INSR_MOUSE	O9w14 mus musculu
22	665	9.7	1382	1 INSR_HUMAN	P06213 homo sapien
23	656	9.6	1297	1 IRR_HUMAN	P14616 homo sapien
24	655	9.6	1477	1 HTK7_HYDAT	Q25197 hydra atten
25	653	9.5	1607	1 MIPR_LYMSI	Q25410 lymsaea sta
26	651.5	9.5	1300	1 IRR_CANPO	P14617 canva porce
27	596	8.7	1367	1 IGR_HUMAN	P08069 homo sapien
28	590	8.6	1390	1 INSR_AEDAE	O93105 aedes aegypt
29	582	8.5	1246	1 INSR_DROME	P09208 drosophila
30	578	8.4	1373	1 IGR_MOUSE	O60751 mus musculu
31	574.5	8.4	1370	1 IGR_RAT	P24062 rattus norv
32	574	8.4	1114	1 RET_HUMAN	P07949 homo sapien
33	571	8.3	987	1 EPB4_HUMAN	P54760 homo sapien

34	569.5	8.3	984	1 EPB1_CHICK	O07494 gallus gall
35	566.5	8.3	984	1 EPB1_RAT	P09759 rattus norv
36	562.5	8.2	977	1 EPB2_MOUSE	O03145 mus musculu
37	560.5	8.2	984	1 EPB1_HUMAN	P54762 homo sapien
38	558	8.1	902	1 EPBB_XENLA	O91736 xenopus lae
39	556.5	8.1	976	1 EPB2_HUMAN	P29317 homo sapien
40	551	8.0	987	1 EPB4_MOUSE	P54761 mus musculu
41	550.5	8.0	985	1 EPBA_XENLA	O91571 xenopus lae
42	544	7.9	1068	1 FAK1_XENLA	O91738 xenopus lae
43	539.5	7.9	1053	1 FAK1_CHICK	O00944 gallus gall
44	537	7.8	757	1 HTL6_HYDAT	P53356 hydra atten
45	532.5	7.8	1052	1 FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

```

RESULT 1
ID   ERB2_HUMAN          STANDARD;          PRT; 1255 AA.
AC   P04626;
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DI   13-AUG-1987 (Rel. 41, Last annotation update)
DE   Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE   (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (tyrosine kinase-type cell
DE   surface receptor HER2) (MIM 19).
GN   ERBB2 OR HER2 OR NGL OR NEU.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX   NCBI_TaxId=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=66118663; PubMed=3003577;
RA   Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA   Saito T., Toyoshima K.;
RT   "Similarity of protein encoded by the human c-erbB-2 gene to
RT   epidermal growth factor receptor.";
RL   Nature 319:230-234(1986).
[2]
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=66070181; PubMed=2999974;
RA   Cousens U., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA   McGee U., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA   Francke U., Levinson A., Ullrich A.;
RT   "Tyrosine kinase receptor with extensive homology to EGF receptor
RT   shares chromosomal location with neu oncogene.";
RL   Science 230:1132-1139(1985).
[3]
RN   [3]
RP   SEQUENCE OF 737-1031 FROM N.A.
RX   MEDLINE=86016729; PubMed=2995967;
RA   Semba K., Kamata N., Toyoshima K., Yamamoto T.;
RT   "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT   c-erbB-1/epidermal growth factor receptor gene and is amplified in a
RT   human salivary gland adenocarcinoma.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
[4]
RN   [4]
RP   VARIANTS VAL-654 AND VAL-655.
RX   MEDLINE=93194196; PubMed=8095488;
RA   Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT   "Characterization of a new allele of the human ERBB2 gene by allele-
RT   specific competition hybridization.";
RL   Genomics 15:426-429(1993).
-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
ALPHA AND AMPHIREGULIN.
-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
(POTENTIAL).
-!- SUBCELLULAR LOCATION: Type I membrane protein.

```

FT	DISULFID	563	576	BY SIMILARITY.			
FT	DISULFID	567	584	BY SIMILARITY. <td></td>			
FT	DISULFID	587	596	BY SIMILARITY. <td></td>			
FT	DISULFID	600	623	BY SIMILARITY. <td></td>			
FT	DISULFID	626	634	BY SIMILARITY. <td></td>			
FT	DISULFID	630	642	BY SIMILARITY. <td></td>			
FT	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY). <td></td>			
FT	MOD_RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY). <td></td>			
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL). <td></td>			
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. . .) (POTENTIAL). <td></td>			
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. . .) (POTENTIAL). <td></td>			
FT	CARBOHYD	259	259	N-LINKED (GLCNAC. . .) (POTENTIAL). <td></td>			
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. . .) (POTENTIAL). <td></td>			
FT	CARBOHYD	571	571	N-LINKED (GLCNAC. . .) (POTENTIAL). <td></td>			
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. . .) (POTENTIAL). <td></td>			
FT	VARIANT	654	654	I -> V. <td></td>			
FT	VARIANT	655	655	/Frid=VAR_004077. <td></td>			
FT	VARIANT	655	655	I -> V. <td></td>			
FT	VARIANT	655	655	/Frid=VAR_004078. <td></td>			
FT	CONFLICT	1170	1170	P -> A (IN REF. 2). <td></td>			
FT	SEQUENCE	1255 AA;	137909 MM;	3939DFDA04DC962 CRC64;			
Query Match							
Best Local Similarity 97.3%;							
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2							
QY	1	MELTALRWMGLLILPRGAS	TGVCTGTDKRLRLPAS	PETHLMDLRLYQGC	GVVQGNL 60		
DB	1	MELTALCRWGLLILPRGAS	TGVCTGTDKRLRLPAS	PEHLMRLHLYQGC	GVVQGNL 60		
QY	61	ELTYLPTNASISFLQDIOE	VOGYVLIAHNO	RVQVPLQRLIRVGTQ	LFEEDNYALAVLNG 120		
DB	61	ELTYLPTNASISFLQDIOE	VOGYVLIAHNO	RVQVPLQRLIRVGTQ	LFEEDNYALAVLNG 120		
QY	121	DLPLNNTPTVTSASGGLRE	LQASLTTELKSGVLIQRN	POCYOITIMKDI	FHNQOLA 180		
DB	121	DLPLNNTPTVTSASGGLRE	LQASLTTELKSGVLIQRN	POCYOITIMKDI	FHNQOLA 180		
QY	121	DLPLNNTPTVTSASGGLRE	LQASLTTELKSGVLIQRN	POCYOITIMKDI	FHNQOLA 180		
DB	121	DLPLNNTPTVTSASGGLRE	LQASLTTELKSGVLIQRN	POCYOITIMKDI	FHNQOLA 180		
QY	181	LTLIDNRSRAHPCHPCSP	MCKSKRCWGESS	EDCQSLTRTVCA	GGCARCKGRLPTDC	CHBQC 240	
DB	181	LTLIDNRSRAHPCHPCSP	MCKSKRCWGESS	EDCQSLTRTVCA	GGCARCKGRLPTDC	CHBQC 240	
QY	181	LTLIDNRSRAHPCHPCSP	MCKSKRCWGESS	EDCQSLTRTVCA	GGCARCKGRLPTDC	CHBQC 240	
DB	181	LTLIDNRSRAHPCHPCSP	MCKSKRCWGESS	EDCQSLTRTVCA	GGCARCKGRLPTDC	CHBQC 240	
QY	241	AAGCTGPRHSGCLCLH	NHSGICEHNPALVTYNT	DFESAPNEGRYTF	EGASCVTACP 300		
DB	241	AAGCTGPRHSGCLCLH	NHSGICEHNPALVTYNT	DFESAPNEGRYTF	EGASCVTACP 300		
QY	241	AAGCTGPRHSGCLCLH	NHSGICEHNPALVTYNT	DFESAPNEGRYTF	EGASCVTACP 300		
DB	241	AAGCTGPRHSGCLCLH	NHSGICEHNPALVTYNT	DFESAPNEGRYTF	EGASCVTACP 300		
QY	301	YNYLSTDVGSCTLYC	PLHNOEVTAE	DGTQRC	EKSKPCARVCYGLG	MOYIKANSKF 360	
DB	301	YNYLSTDVGSCTLYC	PLHNOEVTAE	DGTQRC	EKSKPCARVCYGLG	MOYIKANSKF 360	
QY	301	YNYLSTDVGSCTLYC	PLHNOEVTAE	DGTQRC	EKSKPCARVCYGLG	MOYIKANSKF 360	
DB	301	YNYLSTDVGSCTLYC	PLHNOEVTAE	DGTQRC	EKSKPCARVCYGLG	MOYIKANSKF 360	
QY	361	ELFEGAGCKIKGSLA	FLPESFDGDPAS	TAPLOPOLQV	FEETLEITGVLYI	SAMPDLSLP 420	
DB	361	IQEPRAGCKIKGSLA	FLPESFDGDPAS	TAPLOPOLQV	FEETLEITGVLYI	SAMPDLSLP 420	
QY	361	IQEPRAGCKIKGSLA	FLPESFDGDPAS	TAPLOPOLQV	FEETLEITGVLYI	SAMPDLSLP 420	
DB	361	IQEPRAGCKIKGSLA	FLPESFDGDPAS	TAPLOPOLQV	FEETLEITGVLYI	SAMPDLSLP 420	
QY	421	DLSPVQNLQVIRGRIL	HNHGAVALTQSGIS	MSLGRSLRELG	SGSLALIHNNHLC	CFVHTV 480	
DB	421	DLSPVQNLQVIRGRIL	HNHGAVALTQSGIS	MSLGRSLRELG	SGSLALIHNNHLC	CFVHTV 480	
QY	421	DLSPVQNLQVIRGRIL	HNHGAVALTQSGIS	MSLGRSLRELG	SGSLALIHNNHLC	CFVHTV 480	
DB	421	DLSPVQNLQVIRGRIL	HNHGAVALTQSGIS	MSLGRSLRELG	SGSLALIHNNHLC	CFVHTV 480	
QY	481	PMDOLFRRPHQALLHT	ARPEDECVEG	ELACHOLCAR	HCNGRGPTQ	CNCSQFLRGDGC 540	
DB	481	PMDOLFRRPHQALLHT	ARPEDECVEG	ELACHOLCAR	HCNGRGPTQ	CNCSQFLRGDGC 540	
QY	481	PMDOLFRRPHQALLHT	ARPEDECVEG	ELACHOLCAR	HCNGRGPTQ	CNCSQFLRGDGC 540	
DB	481	PMDOLFRRPHQALLHT	ARPEDECVEG	ELACHOLCAR	HCNGRGPTQ	CNCSQFLRGDGC 540	
QY	541	VEECRVIOGLRE	VEYVNAHCLPCH	DECQPO	NGSVTCFGE	READQVACA	HYKDPFCVABC 600
DB	541	VEECRVIOGLRE	VEYVNAHCLPCH	DECQPO	NGSVTCFGE	READQVACA	HYKDPFCVABC 600

QY 721 FTVSFWLRVPKVSAS-HLETVYKGIWIPGDNVKKIPVATKVLRENTSPKANELLIDEAVV 779
 DB 720 -----LRKRVGLGSAFGTVTKGIWIPGDNVKKIPVATKVLRENTSPKANELLIDEAVV 773
 QY 780 MAGVGSPPYVSRLLIGLITSTVQVLTQVLMFYGCLLDHVRNRRGLSQDLLNMCWQIAKAM 839
 DB 774 MAGVGSPPYVSRLLIGLITSTVQVLTQVLMFYGCLLDHVRNRRGLSQDLLNMCWQIAKAM 833
 QY 840 SYLEDVRLVHRDLAANNVLKSPNHYKITDFGLARLLDIDETEHYHADGKVPILKMALES 899
 DB 834 SYLEDVRLVHRDLAANNVLKSPNHYKITDFGLARLLDIDETEHYHADGKVPILKMALES 893
 QY 900 ILRRFTHQSDVWSYGVTVWELMTFGAKFYDGIIPAREIPDLLEKGERLPPQPICTIDVYM 959
 DB 894 ILRRFTHQSDVWSYGVTVWELMTFGAKFYDGIIPAREIPDLLEKGERLPPQPICTIDVYM 953
 QY 960 IMVKCMWIDSECPRRRELVSSESRMAPDQRFVYIQNEDLGPAFLDSTFYRSLIEDDD 1019
 DB 954 IMVKCMWIDSECPRRRELVSSESRMAPDQRFVYIQNEDLGPAFLDSTFYRSLIEDDD 1013
 QY 1020 MGDLVAEEYLVVQGFPCDDPAPAGAGVWHRHRSSTRSGGDLTLGLPSEEBEAPRS 1079
 DB 1014 MGDLVAEEYLVVQGFPCDDPAPAGAGVWHRHRSSTRSGGDLTLGLPSEEBEAPRS 1073
 QY 1080 PLAPSEAGSDVFDGGLGCAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1139
 DB 1074 PLAPSEAGSDVFDGGLGCAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1133
 QY 1140 SPQPEYVNPQDVRPQPSREGPLPARPAGATLEPAKTLGSKGNVMDVAFGAVEN 1199
 DB 1134 SPQPEYVNPQDVRPQPSREGPLPARPAGATLEPAKTLGSKGNVMDVAFGAVEN 1193
 QY 1200 PEYLTPQGAAPQPHPPAFSPAFLNLYVWDQDPPREGAPPSPTFKGTPTAENPEYLGLDV 1259
 DB 1194 PEYLTPQGAAPQPHPPAFSPAFLNLYVWDQDPPREGAPPSPTFKGTPTAENPEYLGLDV 1253
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 2
 ERB2_RAT STANDARD; PRT: 1257 AA.
 ID ERB2_RAT P06494;
 AC 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).
 GN ERB2 OR NEU.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=86118662; PubMed=3945311;
 RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
 RT "The neu oncogene encodes an epidermal growth factor receptor-related protein."
 RT Nature 319:226-230(1986).
 RL Nature 319:226-230(1986).
 RP SEQUENCE OF 852-905 FROM N.A.
 RC TISSUE=Sciatic nerve;
 RX MEDLINE=91222560; PubMed=2025425;
 RA Lai C., Lemke G.;
 RT "An extended family of protein-tyrosine kinase genes differentially
 RT expressed in the vertebrate nervous system."
 RT Neuron 6:691-704(1991).
 RN [3]
 RP STRUCTURE BY NMR OF 650-668.
 RX MEDLINE=92155181; PubMed=1346763;
 RA Gullick W.J., Bottomley A.C., Lofis F.J., Doak D.G., Mulvey D.,
 RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
 RT "Three dimensional structure of the transmembrane region of the proto-
 RT oncogenic and oncogenic forms of the neu protein."
 RL EMBL J. 11:43-48(1992).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PIR: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X03362; CAZ27059.1; ALT_INIT.
 CC PIR: A24362; TVRTNU.
 CC HSSE: P11362; IFGK.
 CC InterPro: IPR000494; EGFR_L_domain.
 CC InterPro: IPR000719; Euk_Pkinase.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR001245; Tyr_Pkinase.
 CC InterPro: IPR004019; YLP_motif.
 CC Pfam: PF00069; Pkinase; 1.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF01030; Recep_L_domain; 2.
 CC Pfam: PF02757; YLP; 2.
 CC ProDom: PD000001; Euk_Pkinase; 1.
 CC SMART: SM00261; FU; 3.
 CC SMART: SM00219; TYKIC; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC Proto-oncogene; Disease mutation.
 CC SIGNAL 1 21 POTENTIAL.
 CC CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 CC DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 655 677 POTENTIAL.
 CC DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
 CC DOMAIN 159 369 CYS-RICH.
 CC DOMAIN 473 646 PROTEIN KINASE.
 CC NP_BIND 722 989 ATP (BY SIMILARITY).
 CC BINDING 728 736 ATP (BY SIMILARITY).
 CC ACT_SITE 755 755
 CC ACT_SITE 847 847
 CC DISULFID 196 205 BY SIMILARITY.
 CC DISULFID 200 213 BY SIMILARITY.
 CC DISULFID 221 228 BY SIMILARITY.
 CC DISULFID 225 236 BY SIMILARITY.
 CC DISULFID 237 243 BY SIMILARITY.
 CC DISULFID 241 253 BY SIMILARITY.
 CC DISULFID 256 265 BY SIMILARITY.
 CC DISULFID 269 296 BY SIMILARITY.
 CC DISULFID 300 312 BY SIMILARITY.
 CC DISULFID 316 332 BY SIMILARITY.
 CC DISULFID 335 339 BY SIMILARITY.
 CC DISULFID 513 522 BY SIMILARITY.

FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.
 FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.4%; Score 5850; DB 1; Length 1257;
 Best Local Similarity 85.8%; Pred. No. 3.2e-306; Indels 10; Gaps 4;
 Matches 1084; Conservative 52; Mismatches 118;

QY 1 MEALALCRMGILLALPPGAASTOYCTGTDMLRLPASPEHLMDLRHLYOGQVVOGML 60
 DB 1 MELALACRMGFLALLPPIAGTGTCTGTDMLRLPASPEHLMDLRHLYOGQVVOGML 60
 QY 61 ELTYLPNALSIFQDIQEVGYVLIHNOVROVPLORLIVAGTQLFEDVNAVLNDNG 120
 DB 61 ELTYVPANASLIFQDIQEVGYVLIHNOVKRVPLORLIVAGTQLFEDVNAVLNDNR 120
 QY 121 DPLNNTTPVT-GASPGGLRELOLRLSLTEILKGVLIORNPOLCYODTILMKDIFHKNNQL 179
 DB 121 DPQDNVASTIRGRTPEGLRELQRLSLTEILKGVLIORNPOLCYODTILMKDIFHKNNQL 180
 QY 180 ALTLIDTNRSPRACPCSPWCKGSRGWESSEDCQSLTRIVACAGCARCKGPLETDCHEQ 239
 DB 181 APVDIDTNRSPRACPCSPWCKGSRGWESSEDCQSLTRIVACAGCARCKGPLETDCHEQ 240
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNDTESNPNEGRTFEGASCVTAC 299
 DB 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNDTESNPNEGRTFEGASCVTAC 300
 QY 300 PNYLSTDVSGCTLVCLHNOEVTAEEDGTQRCCKSPCARVCYGLGMQYIKANSKFTGI 359
 DB 301 PNYLSTEVGSCCTLVCPNNQOEVTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFTGI 360
 QY 360 TELPACCKKIFGSLAPSPGDDPASNTPAPIOPEDLOVETLEETITGYLISAMPDSL 419
 DB 361 NMOEFDCCKKIFGSLAPSPGDDPASNTPAPIOPEDLOVETLEETITGYLISAMPDSL 420
 QY 420 PDLSPFONLOVIRIRLIHNGAYSLTLOGLISWLGISLRELSGALIHNNHLCFVHT 479
 DB 421 RDLSPFONLRIIRIRLIHNGAYSLTLOGLISWLGISLRELSGALIHNNHLCFVHT 480
 QY 480 VPMDQCFERNPHOALIHANRPEDE-CYGBGLACHQICARHCHGSPGTQCVNCSOTLRQ 538
 DB 481 VPMDQCFERNPHOALIHANRPEDE-CYGBGLACHQICARHCHGSPGTQCVNCSOTLRQ 540
 QY 539 ECVEECRVLQGLPREYVNAHCLPCHECOPONGSVTCFPEADQCAAHYDPPFCYA 598
 DB 541 ECVEECRVMKGLPREYVSDKRCPCHECOPONGSVTCFPEADQCAAHYDPPFCYA 600
 QY 599 RCBGKVPDLSTYPMIKWFPDEEGACCPINCHSCVDLDDKCGPAORASPLTSVAY 658
 DB 601 RCBGKVPDLSTYPMIKWFPDEEGACCPINCHSCVDLDDKCGPAORASPLTSVAY 660
 QY 659 VGILLVVLGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAOMRIKEF 718
 DB 661 VGILLVVLGVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPQAOMRIKEF 720
 QY 719 ANFIVSFMLVAPKYSAS-HLETYKGIWIDGENVKIPVAIKYLRENTSFKANKELIDEA 777

DB 721 E-----LRKYKVLGSAFGTIVKGIWIDGENVKIPVAIKYLRENTSFKANKELIDEA 773
 QY 778 YMGAGVSPYRSLRIGLCLTSTVQVLTQMPYGLDVRHNRRLSGQDLNMCQIAK 837
 DB 774 YMGAGVSPYRSLRIGLCLTSTVQVLTQMPYGLDVRHNRRLSGQDLNMCQIAK 833
 QY 838 GMSYEDVRLVHRDLAANVLYKSPNRYKLTDFGLARLLDIDETRYHADGKVPKMMAL 897
 DB 834 GMSYEDVRLVHRDLAANVLYKSPNRYKLTDFGLARLLDIDETRYHADGKVPKMMAL 893
 QY 898 ESILRRFTHSDVMSYVTVWELMTFCAPKYDGIIPAREIPDLLEKGERLPQPICTIDV 957
 DB 894 ESILRRFTHSDVMSYVTVWELMTFCAPKYDGIIPAREIPDLLEKGERLPQPICTIDV 953
 QY 958 YMIWKCMMIDSECPREFRELVSERFARMARDPQRFVIONEDLGPASPLDSTFYSRLSD 1017
 DB 954 YMIWKCMMIDSECPREFRELVSERFARMARDPQRFVIONEDLGPASPLDSTFYSRLSD 1013
 QY 1018 DDMGDLVDAEFLYVQGGFCPPDPAPAGGVHHRSSSTRSGGDLTLGLEPSEEEAP 1077
 DB 1014 DDMGDLVDAEFLYVQGGFCPPDPAPAGGVHHRSSSTRSGGDLTLGLEPSEEEAP 1073
 QY 1078 RSLPAPSEAGSDVFDGDLGCAAKGLQSLPTHDPSPLOYSSEDTVPLPESTDGYVAPL 1137
 DB 1074 RSLPAPSEAGSDVFDGDLGCAAKGLQSLPTHDPSPLOYSSEDTVPLPESTDGYVAPL 1133
 QY 1138 TCSPOPEYVNOVDVAPQPPSPREGPLAPARAGATLEPAKTLSPCKNVVVDVAFAGAV 1197
 DB 1134 ACSPOPEYVNOVDVAPQPPSPREGPLAPARAGATLEPAKTLSPCKNVVVDVAFAGAV 1193
 QY 1198 ENPEYLTFRQGAAPRHPFAPSPAFDMLYVWDQPPRGAAPSTFTKPTAENEYIGL 1257
 DB 1194 ENPEYLTFRQGAAPRHPFAPSPAFDMLYVWDQPPRGAAPSTFTKPTAENEYIGL 1253
 QY 1258 DVPV 1261
 DB 1254 DVPV 1257
 RESULT 3
 ERB2_MESAU STANDARD; PRT: 1254 AA.
 ID ERB2_MESAU
 AC Q60553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
 GN ERB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxId=10036;
 RN [1] _TaxId=10036;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RT Cloning and activation of the Syrian hamster neu proto-oncogene.";
 RL Gene 140:251-255(1994).
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D1295; BAA03801.1; -
 DR HSP: P11362; IEGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Dkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; Fu; 3.
 DR SMART: SM00219; TykC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 1254 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 158 368 CYS-RICH.
 FT DOMAIN 472 644 CYS-RICH.
 FT DOMAIN 720 987 PROTEIN KINASE.
 FT NP_BIND 726 734 ATP (BY SIMILARITY).
 FT BINDING 753 753 ATP (BY SIMILARITY).
 FT ACT_SITE 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 268 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.
 FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 587 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 658 658 V -> E (IN ONCOGENIC NEU).
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
 FT SEQUENCE 1254 AA: 138252 MW: 974C3791C21F2BB1 CAC64;

Query Match 85.1%; Score 5834.5; DB 1; Length 1254;
 Best Local Similarity 85.3%; Fred. No. 2.2e-305;
 Matches 1077; Conservative 61; Mismatches 115; Indels 9; Gaps 3;
 1 MELALCFRWGLLALPPGAATQVCTGDMRLRPAPETHLDMLRHLVQCCQVQGNL 60
 1 MELAMCGGLLALSPASSTQVCTGDMRLRPAPETHLDMLRHLVQCCQVQGNL 60
 61 ELTYPTNALSFLQDIOEVGYVLIANQVQVPLQRLRYRGITQLEEDNALVLDNG 120
 61 ELTYPANATLSFLQDIOEVQGYMLIAHQVAVPLQRLRYRGITQLEEDKALAVLDNR 120
 121 DELNNTTPVTGASPGGLRELQRLSTELIKGVLLQRPOLCYODTILMKDIFHNQOLA 180
 121 DPLDNTTATGTPEGLPELQRLSTELIKGVLLQRPOLCYODTILMKDIFHNQOLA 180
 181 LTLIDNRSRACHPSPCMKGSRCWGESSEDCQSLTRTVAGGACARCKPLPTDCHEQC 240
 181 PVDIDNRSRACHPSPCAPACKDNHCWASPEDCQTLTGITAPRAVPAARARLPTDCHEQC 240
 241 AAGCTGPKKSDCLALHFNHSGICELHCPALVTYTDTPESMPNPEGRTFGASCVTCP 300
 241 AAGCTGPKKSDCLALHFNHSGICELHCPALVTYTDTPESMPNPEGRTFGASCVTCP 300
 301 YNYLSTDVGSCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLMWYIKANSKFIGIT 360
 301 YNYLSTEVGSCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLMWYIKANSKFIGIT 360
 361 ELFFAGCKKIFGSLAFLESFGDPASTAPLQPOLQVFEITLGYLYSAMPDCLP 420
 361 IDEFAGCKKIFGSLAFLESFGDPASTAPLQPOLQVFEITLGYLYSAMPDCLP 420
 421 DLSVQNLQVIRGRIHNGAYSLTQGLISWLGRLSRLRELSGALVHNNHLCFPHV 480
 421 DLSVQNLQVIRGRIHNGAYSLTQGLISWLGRLSRLRELSGALVHNNHLCFPHV 480
 481 PVDQLFRRNHQALHTANRPEDECVGEGIALCHQLCARGHCWGPPTCCVNCQPLRGEC 540
 481 PVDQLFRRNHQALHTANRPEDECVGEGIALCHQLCARGHCWGPPTCCVNCQPLRGEC 540
 541 VEECVLQGLPREYVNAHCHLPCHECQPNQNSVTCFPEADQCACAHYKPPCVNAC 600
 541 VEECVLQGLPREYVNAHCHLPCHECQPNQNSVTCFPEADQCACAHYKPPCVNAC 600
 601 PEGVAPDLSTYMPIMWKPDEBEGACQPCPINTCHSCVDLDDKGPACQBRASPLTSIYAVG 660
 601 PEGVAPDLSTYMPIMWKPDEBEGACQPCPINTCHSCVDLDDKGPACQBRASPLTSIYAVG 660
 661 ILVVVVLGVVFGILLKRRQOKTRKTYMRLLOETELVBPPLPSGAMPNQAOKRIKEFN 720
 661 ILVVVVLGVVFGILLKRRQOKTRKTYMRLLOETELVBPPLPSGAMPNQAOKRIKEFN 720
 721 FTVSFWLRPKVSAS-HLETYKGIWIPDEGNVKIPVAIKVLRNTSPKANKEIDVAV 779
 721 FTVSFWLRPKVSAS-HLETYKGIWIPDEGNVKIPVAIKVLRNTSPKANKEIDVAV 779
 774 MAGLSPYASRLGLCTSTVQLVQLMPYGGLDHVEHGRGLSQDLINWCQIANGM 833
 774 MAGLSPYASRLGLCTSTVQLVQLMPYGGLDHVEHGRGLSQDLINWCQIANGM 833
 840 SYLEQVRLVHRDLAARNVLVSPNFKTTDGLARLLIDDEFYVADGKVPKMALES 899
 840 SYLEQVRLVHRDLAARNVLVSPNFKTTDGLARLLIDDEFYVADGKVPKMALES 899
 894 ILRRFTHQSDWVSQVTVWELMTFGAPYDGIAPREIPDLLEKGERLPQPPICITDYM 953
 894 ILRRFTHQSDWVSQVTVWELMTFGAPYDGIAPREIPDLLEKGERLPQPPICITDYM 953
 954 IMVKCMMDSECRPRELVSEFSMAADPQRFVIVIQNEDIGPASPLDSTYRSLTEDD 1013
 954 IMVKCMMDSECRPRELVSEFSMAADPQRFVIVIQNEDIGPASPLDSTYRSLTEDD 1013

QY 1020 MGDLVDAEYLVPOGPFCDPAPGAGMGNHRRSSSTRSGGDLTGLPSEEPAPRS 1079
 DB 1014 MGDLVDAEYLVPOGPFCDPAPGAGMGNHRRSSSTRSGGDLTGLPSEEPAPRS 1073
 QY 1080 PLAPSEGAGSDVDFDGLGMAAKGLOSLPTHDPSLQRYSEDPVPPVLPSETDGVAPLTC 1139
 DB 1074 PLAPSEGAGSDVDFEELGMGATKQGSISPRDLSPLOKRYSDPLPLPTETDGVAPLAC 1133
 QY 1140 SPOPEYVNPDPVRPQPPSPRPGPPLPAPACATLIERAKTSLPGXGVYKDYFAFGAVEN 1199
 DB 1134 SPOPEYVNPDPVRPQPPSPRPGPPLPAPACATLIERAKTSLPGXGVYKDYFAFGAVEN 1193
 QY 1200 PEVLTPOGGAAPQPPPPAPFSPAPDNLYYWDQDPPRCAPSTFKGTPTAENPEYLGIDV 1259
 DB 1194 PEVLTPOGGAAPQPPPPAPFSPAPDNLYYWDQDPPRCAPSTFKGTPTAENPEYLGIDV 1252
 QY 1260 PV 1261
 DB 1253 PV 1254
 RESULT 4
 EGFR_HUMAN STANDARD; PRT; 1210 AA.
 ID EGFR_HUMAN
 AC P00533; P06268; Q14225; Q9CMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
 AC O00688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CX NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells";
 RT Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=97076686; PubMed=8918811;
 RA Reiter J.L., Mahle N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RX MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGR) in ovarian cancer";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX TISSUE=Placenta;
 RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Scheel Sinclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balaubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
 RA Mahle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Scheel C.M.,
 RA Lampland A.L., Balaubramaniam S., Crossley T.O., Magnuson T.R.,
 RA Mahle N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBS databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krutiger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Symmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells";
 RL Biochem. Biophys. Res. Commun. 124:1125-1132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kitchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription";
 RL Oncogene Res. 1:375-386(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and supercoiled DNA." ;
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Howk R., Glyvol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor." ;
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts." ;
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Sato K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor." ;
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=9825196; PubMed=9556602;
 RA Abe Y., Odaoka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor." ;
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens." ;
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp10 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TBGR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X00588; CAA25240.1; -
 CC DR EMBL; U95089; AAB53063.1; -
 CC DR EMBL; U48722; AAC50802.1; -

DR EMBL; U48723; AAC50804.1; -
 DR EMBL; U48724; AAC50796.1; -
 DR EMBL; U48725; AAC50797.1; -
 DR EMBL; U48726; AAC50798.1; -

Query Match 44.7% Score 3063; DB 1; Length 1210;
 Best Local Similarity 48.7%; Pred. No. 7e-157;
 Matches 620; Conservative 178; Mismatches 360; Indels 114; Gaps 24;

11 LLLALLPFGAA--STVCTGTDMLRLRLASPTHTLDMRLHLYQCGVYQGNLELYLPN 68
 14 LLAALCPASRLAEKKVQGSNKLTLQGTEDHFLSLQRMFNNCEVVLGNLEITYVGRN 73
 69 ASLSFLQDIOEYGVYLIAHQVROVPLORIRIYVGTOLFEDNVALAVLNDPPLNNTTP 128
 74 YDLSLTKIQAAGVAVLALNTVERIPLENQIRGNMYNSYALAVLND----- 126
 129 VTGASPGGLRELQLSLTEILIKGVLIQRNPQLCYQDTILMKDIFKNNQALALIDTNR 188
 127 ---ANKTLKELPMENLOEILHGAVERFSNNPALCVESIQWRDIVSDFLSNMSMDFONH 183
 189 SRACHPGSPMGKSGRCWGESSEDDQSLRTVCAAGCA-RCKGFLPTDCHEQCAAGCTGP 247
 184 LOSCKCPSPCPNGSCWAGBNCKLTKITLCAQCCSRCKSPSCCHNOCAAGCTGP 243
 248 KHSDDLACIHNHSGICELHCPALVTYNTDFFESNPNBGRYTFGASCVTACPYNLSTD 307
 244 RESDLVCRKFRDEATCKDTCPLMLYPTTYQMDVNEGKYSFGATCVKCCPRVYVTD 303
 308 VQSCITVCPLNQETATDGTORCEKSKPCARVCYGLGMQYIKANSFGLTTELE-FAG 366
 304 HGSVCVACGADSYEN-EEGVKCKCKGPRKVCNNGIGIEFK-DSLSINATNFKHKN 361
 367 CKKIFGSLAFIPESHGDPASNTAPLOEQQLQVETLEETLEIGYLYISAMPDPLPLSFQ 426
 362 CTSISGDHLIPVAFRGDSFHTPTPLDQELDLKTYEITGFLILQMWENRDLDAFE 421
 427 NLQVTRGILHNHAYSLTLQGLISWGLRLRLRELGSGLALIHNTLFCYHYVWDLF 486
 422 NLEIRGRTKOHGFSLAVSLNITSLRLSLKEISDDVILISGNKNCVANTIMWKLFL 481
 487 RNPQALLHTANRPDECVGEGGLACHQICAGHCKMGPPTQCVNCSOLFROECVEEGRV 546
 482 GTSQKTKIISNGENSKATGOVCHALCSPEGCGPFRDVCGRNRSRBECDKNL 541
 547 LQGLPREVYNARHCLPCHPEQPNQSVTCFPEADQCVACAHYKDPFCVACRPSGVKP 606
 542 LEGEPRFEVENSECICQHPCECLPQAMNITCTGRPDNOCICAHYIDGHCVKTCGAVMG 601
 607 DLSYPIWKFPEDEGACQPCINCTHSCVDDDDGCAPEQAASPLTSIVSAVVG---ILL 663
 602 ENNTL-VKTYADAGHVCHLCPNCTYGTGGLGCPNNGKIP--STATMVALLLL 658
 664 VVVLGVVFGILIKERQCKIRKYTRRLLOETELVEPLTPSGAMPNOAMRLIKEFNNTV 723
 659 VVALGIG---LPMRRHIVKRTLRLLQHELVEPLTPSEARNQALRLIKE----- 709
 724 SFMLRVPVNSHSETYTKGIWPDGENVKIPVAILKRLRENTSPRANKIIDEAYVAVG 783
 710 TEFKIKLIGSAGATYKGLMPEGEVYKIPVAILKRLRENTSPRANKIIDEAYVAVS 769
 784 GSPYVSRLLGICLSTVQVLTQMPYCGILDHVENNGRLGSGOOLMCMQIAGMAYLE 843
 770 DNPVHCLLGLICLSTVQVLTQMPYCGILDHVENNGRLGSGOOLMCMQIAGMAYLE 829
 844 DVRLVHRDLAARVLYVSPNHVKTITDGLARLLDIDETEHADGKVPKIMMALESILRR 903
 830 DRLVHRDLAARVLYVTPGHVKTITDGLAKLLGAEEKEYNAEGKVPKIMMALESILRR 889
 904 RFTQSDVMSYGVTVVEMLMFGAKPYDGIAREIPDLLENGERLPQEPICITDVMYIMVK 963
 890 IYTHQSDVMSYGVTVVEMLMFGAKPYDGIAREIPDLLENGERLPQEPICITDVMYIMVK 949

QY 964 CWMIDSECRPRELVESESRMARDPQRFVIO-NEDLGPASPLDSTFYRSLIEDDDMGD 1022
 DB 950 CWMIDADSRKPELLEISKMKAPDPRVYVIGDERMLPSTDSNFTALMDERDMD 1009
 QY 1023 LVDAEYLVEQGFPCDPAPGAGVHHRHRSSTRSGGGLTLGLPSEERAPRPLA 1082
 DB 1010 VVDAEYLVEQGFPCDPAPGAGVHHRHRSSTRSGGGLTLGLPSEERAPRPLA 1035
 QY 1083 PSEGAGSDVFDGGLKMAAGLQSLPHDPSLQRYSEDPVLPSET--DGVAFLTGS 1140
 DB 1036 SSSLSATSN--NSTVACIDRNGLOSCPKEDSFLOQYSSDPTGALTEDS:IDTFL----- 1087
 QY 1141 POPEYVQPDVRFQPSPREGPLPAARPGATLERAKTLSPGKNGVYKDVFAFGAVENP 1200
 DB 1088 FVEPEYINQ-SVPRKPRAGSVQNPVYHNGPLNP-----APSRDPHYQD--PHSTAVGNP 1136
 QY 1201 EYL-TPQGGAAPCPHPPPAPSPAFDNLVYWDQ-----DP-----PERGAPPSFF 1243
 DB 1137 EYLVNTVQ-----FTCVNSTFDSFAMWAKGSHQISLNDPDYQODFFPKAKENGIF 1187
 QY 1244 KGTPTAENPEYL 1255
 DB 1188 KGS-TAENAEYL 1198

RESULT 5
 EGFR_MOUSE
 ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
 AC 001279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RC STRAIN=BALB/C; TISSUE=Liver;
 RX MEDLINE=93026370; Pubmed=1408137;
 RA Avivi A., Skorecki K., Yavon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 (bek/KGFR) gene.";
 RL Oncogene 7:1957-1962(1992).
 RN (2)
 RC STRAIN=BALB/C, and CD-1; TISSUE=Liver, and Decidua;
 RX MEDLINE=93126380; Pubmed=7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 in mouse blastocysts during delayed implantation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN (3)
 RC STRAIN=BALB/C; TISSUE=Liver;
 RX Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
 RA Hibbs M.L.;
 RL [4]
 RC STRAIN=B6/C3; TISSUE=Liver;
 RX MEDLINE=94170986; Pubmed=6125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 receptor tyrosine kinase.";
 RL Genes Dev. 8:339-413(1994).
 RN (5)
 RC TISSUE=Brain;
 RX MEDLINE=91232866; Pubmed=2030916;
 RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

RT "Comparison of EGF receptor sequences as a guide to study the ligand
 RT binding site.";
 RL Oncogene 6:673-676(1991).
 RN (6)
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN=C3H;
 RA Bissinger D.P., Serrero G.;
 RT Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X78987; CAA5587.1; -
 DR EMBL: U03425; AAA17899.1; -
 DR EMBL: X59698; CAA42219.1; -
 DR EMBL: L06864; AAA53029.1; -
 DR EMBL: Z12608; CAA78249.1; -
 DR HSSP: P11362; 1PGK.
 DR MGD: MGI:95294; Egfr.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SMO0261; Fu; 3.
 DR SMART: SMO0219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 647
 FT TRANSMEM 648 670
 FT DOMAIN 671 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT DOMAIN 1028 1071
 FT DOMAIN 714 981
 FT NP_BIND 720 728
 FT BINDING 747 747
 FT ACT_SITE 839 839
 FT DISULFID 190 199
 FT DISULFID 194 207
 FT DISULFID 215 223
 FT DISULFID 219 231
 FT DISULFID 232 240
 FT DISULFID 236 248
 FT DISULFID 251 260
 FT DISULFID 264 291
 FT DISULFID 295 307
 FT DISULFID 311 326
 FT DISULFID 329 333
 FT DISULFID 333

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF222F5 CRC64;

```

Query Match 44.3%; Score 3037; DB 1; Length 1210;
 Best Local Similarity 48.6%; Pred. No. 1,7e-155;
 Matches 622; Conservative 172; Mismatches 367; Indels 118; Gaps 26;

```

QY 11 LLLALLPPGAA--STOYCTGTDMKRLPASPETHLMDLRHLYOCQCVQVQGNLETTYPTN 68
DB 14 LTRLLCAAGALEEKVCQCGISNRNLTLQGTEDHFLSLQRYNNCEVVLGNLETTYQRN 73
QY 69 ASLSFLQDIOEVQYVLIANQVROVPLQRLIRVQTLQFEDNYALAVLDNGPPLNNTTP 128
DB 74 YDSFLKTIQEVAVYLIANTVERIPILENLQIRGNALYENTYALALSN----- 124
QY 129 VTGASPGGLRELOLRSLTELKXGVLLQBNPOLQYOTLIIMKDI-----FKNNQALTLI 164
DB 125 -YGTNRKGLRELKRLNLOELILGAVRSNNPILCNMDLTQWRDVLQNVFSSNMSMDL--- 180
QY 185 DTNRSRACHPCSPKSGSRCKGSESSDQSLTRTVACAGCA-RCKGPLPTDCHEQCAAG 243
DB 181 -QSHPSSCPCKDPCSPKSGCWGGGEENCQKLTIKILCAQCSHRGRSPSDCHNQCAAG 239
QY 244 CTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPPEERYTFGACVYACPNY 303
DB 240 CTGERESDCLVCCQFQDEATCTKDCPPLMYNPTTYMDVNPBGKVSFGATCVKCPERNY 259
QY 304 LSTDVSGCTLVCPILHNOEYTAEDGTQRCCKSKCAVCGYGLQMOVKASKRTGITELE 363
DB 300 VMTDHGCVACCGDYEV--EEDGIRKCKKODGCRVYVNGIGIGERK-DTLSINATNIK 357
QY 364 -FAGCKKIFGSLAFSPFQDPAKSNAPLQPELOQVFTLEBTITGLVYISAMPDSLPLD 422
DB 358 HFKYCTAISGLHLPLVAFKGDSEFTRPPLDPRLELTIKVKEITGFLLLQAMPDMWTDL 417
QY 423 SVEPNQOVIRGRILHNGAVSLTQGLISWGLRSLELSSGLALIHNNTHLFCVHTPV 482
DB 418 HAFENLEIIRGRTHQHGQFSLAVGLNITSLGLRSLEISDGVYIISGNRNLCTYANTIM 477
QY 483 DQLFRRNHQALLHTANPEDECEVGEGLACHQLCARHGCMWGPPTQCVNCSQFLRGCEVE 542
DB 478 KKLFGTNGQTKTKIMNNAEKDCAVNVHCNPLSSSECCWGPPEPDDCVSCNVSRGREVE 537
QY 543 ECRVLQGLPREYVVARHCLPCHPECQFONGSVTCFGEADQCAVCAHYKPPPCVARGCS 602

```

```

DB 538 KNILGEGREFENSECICQHCPCILPQAMNITCTGRGDNCCQCHAYIDGPRCTVTCRA 597
QY 603 GYKPDLSYPIKWEDEBACQPCPINCTHSCVDLDDKCPAPQASPLTSVSAVGL 662
DB 598 GINGENNNTL-VWKYADANNVCHLCHANCTYGCAGPGLQCEVWPSGPKIPSIATGVGL 656
QY 663 LVVVLGVVGI-LIKRQOKIRRYTMRRLLOETELVEPLTPSGAMPNQOMRLKXPNNF 721
DB 657 LFLVV-VAGIGLPMRRRIIVKRTIRLLORELEVEPLTPSGEAPNOHLRLKE---- 711
QY 722 TVSFMLRVKVSASHLETYYKGIWIPDGENVKI PVAIKYLRENTSFKANKELDEAYVNA 781
DB 712 --TEFKIKVLAGAGATGYKGIWIPDGENVKI PVAIKELREATSKANKELDEAYVNA 769
QY 782 GVSFVPSVSLGLCLTSTVQVLTQMLPFGCLLDHREANGRLASQDILLWCNOIAGMSY 841
DB 770 SYDNPFVCRLLGLCLTSTVQVLTQMLPFGCLLDHREANGRLASQDILLWCNOIAGMSY 829
QY 842 LEDVRLVHRDLAARNVLPVSPNHVKITDPGLRLDIDETEVHADQKVPIMMALESIL 901
DB 830 LEDRRLVHRDLAARNVLPVSPNHVKITDPGLRLDIDETEVHADQKVPIMMALESIL 889
QY 902 RRRFTQSDVMSYGVTVWELMTFGAKPYGIPARSLIPDLLEKGERLPQPPCTIDVYIM 961
DB 890 HRYTHQSDVMSYGVTVWELMTFGAKPYGIPARSLIPDLLEKGERLPQPPCTIDVYIM 949
QY 962 VKCMMDISCRPPRELTVSEFSRMAADPQRFVIO-NEDLGPASPLDSTFFYSILEDDM 1020
DB 950 VKCMMDIADSRKFRRLIIEFSKMAADPQRFVIO-NEDLGPASPLDSTFFYSILEDDM 1009
QY 1021 GDLVDAEYLVPQGFCEPDPAPGAGVYHHRSSSTRSGGDLTLGLEPSEEBEAPSP 1080
DB 1010 EDVVDADDEVILTPQGFCEPDPAPGAGVYHHRSSSTRSGGDLTLGLEPSEEBEAPSP 1035
QY 1081 LAPSEAGSDVDGDLGMAAGKQSLPFDHDSPLQRYSEDPVLPSPST--DGYAPRLT 1138
DB 1036 LSSLSATSN---NSTVACINRNGSCRYKEDAFQRYSSDPGATETDNIDAFV--- 1087
QY 1139 CSPOEYVQVQDVPRQPPSPRGPPLPAAPAGATLERAKTSLPGKGVVKKDYAFAGAVE 1198
DB 1088 --PVPEYVQV-SVPRKPAQVQNPVYHNGPLRP-----APGRHLQHN--PHSNAG 1134
QY 1139 NPEYLT-TPQGAAPQHPPPASPAFNUITYDQ-----DP-----PERGAPS 1241
DB 1135 NPEYLTAAQ-----PTCLSSGSPALWIQKSHQMSLDNDYQODPFPKETKPN 1185
QY 1242 TFGATPAENPEYLGIDVP 1260
DB 1186 IFKG-PTAENABYLKRAP 1203

```

RESULT 6
 ERB4 HUMAN STANDARD; PRT: 1308 AA.
 ID ERB4 HUMAN
 AC 015303;
 DT 15-DEC-1998 (rel. 37, last sequence update)
 DT 15-JUN-2002 (rel. 41, last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast Carcinoma;
 RX MEDLINE=93189574; PubMed=8383326;
 RA PLOWMAN G.D., CULOUSCOU J.-M., WHITNEY G.S., GREEN J.M., CARLTON G.W.,
 RA FOY L., NEUBAUER M.G., SHOYAB W.,
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

RT epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester.";
 RL J. Biol. Chem. 272:26761-26768(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
 CC RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
 CC LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: L07868; AAB59446.1; -
 CC DR HSSP: P11362; 1FGK.
 CC DR Genew: HGNC:3432; ERBB4.
 CC DR MIM: 600543; -
 CC
 CC InterPro: IPR000494; EGFR_L_domain.
 CC InterPro: IPR000719; Euk_Pkinase.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR001245; Tyr_Pkinase.
 CC InterPro: IPR004019; YLP_motif.
 CC Pfam: PF00069; Pkinase; 1.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF01030; Recep_L_domain; 2.
 CC Pfam: PF0257; YLP_2.
 CC Prodom: PD000001; Euk_Pkinase; 1.
 CC SMART: SMO0261; FU; 4.
 CC SMART: SMO0219; Tyrc; 1.
 CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 CC PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 CC PROSITE: PS50011; PROTEIN KINASE DOM; 1.
 CC Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 CC Alternative splicing.
 CC KW SIGNAL 1 25
 CC FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 CC FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 CC FT TRASMEN 652 675 POTENTIAL.
 CC FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
 CC FT DOMAIN 186 334 CYS-RICH.
 CC FT DOMAIN 496 633 CYS-RICH.
 CC FT DOMAIN 718 985 PROTEIN KINASE.

FT NP_BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 269 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 350 352 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1189 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPIC 626 648 NGPTSDCIYVPTGTSTPQHA -> ISSIEDCIGLMD
 FT (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA, 146807 MW, 54AAB80985D86761 CRC64;
 Query Match 42.1%; Score 2864.5; DB 1; Length 1308;
 Best Local Similarity 44.2%; Pred. No. 2,8e-147;
 Matches 598; Conservative 185; Mismatches 390; Indels 179; Gaps 29;
 QY 9 WGLLLALLPRGA-----STQVCTGTMTKRLRPSPELTHDMLRHLTYQGGVQVQGNLELY 64
 DB 8 WTVSLVLAAGTVQPSDSQVCAGTGNKLSLSDEQYRALAKYENGEVWGNLEITS 67
 QY 65 LPTNASTLPLQDIOEVQGVLLAHNGVQVPLQRLRIVRGTOLEFEDNYALAVLDNGDPLN 124
 DB 68 IENRDLISFLRSVREYTVGLVALNGFVPLENLIIRGTLYEPRYALAIPLNYRKDG 127
 QY 125 NTTPVYTGASRGRLQLNSTELLKGVLLIQNPOLCYODITLMDIDHKNQALLTLI 184
 DB 128 NF-----GLQELGNKLTLELNGVYVDNKLCTADITHWDIVRNMPSPSLTLV 178
 QY 185 DTRSPRCHPCSPMGKSRWCSESSDQSLRTVQAGG-AACKSPPLPTDCCHQCAAG 243
 DB 179 STNGSGCGRCHNSCTG-RWCGPTENHCQTLRTVCAEQDGRGCTPYVSDCCRRCAAG 237
 QY 244 CTGPKHSDCIACHFNHSGICELHCAALVTYNTDTRESMNPBGRKTFGASCTYACPYNY 303
 DB 238 CSBPKQTDPCFACNPFNDSGACTVQCQTFVYNTTQLNEHNAKTYGAFVCKKCPHNF 297
 QY 304 LSTDVSGCTLVCPHNOEVTABDGTORCEKSPCARVCGYGLMQYIKANSKFIQITBLE 363
 DB 298 V-VDSSSCVACAPSSKMEV-ENNGIMKRCCTDICKAODGIOTGSLMSAQTVSSNIDK 355
 QY 364 FAGCKIFGSLATPSPFDDDPASNTAPLOPELOVPELTLEITGLVYISAWDSLPDLS 423

```

Db      356 FINCTKINGNLIFLVTGHCDDPYNAIEADPEKXNVERTREITGFLNTQSPMNMDFS 415
Qy      424 VEQNLQYIRGRILHNGAVSLTLOGLGSLWLBRELSGGLAIHHTHLCVHTYPMW 483
Db      416 VFSNLVTIGGRVLSGLSLTLKQOGLTSLQFOSLKEISGNYITDNLSCYHTHMT 475
Qy      484 QLEFNPHQALLHTANREDECEVCEGLACHOLCARGHCMGPGPTCVNCSQFLRGOECVEE 543
Db      476 TLFSTINQRIYIRDNRAENCTAREGWCNHLCSDDGCGWGPDDCCSCRFSGRICTES 535
Qy      544 CRVLOGLPREVYANRHLCPHRECPQ-QNGSYTCFGEBAQCAVCAHYKPPCVACPS 602
Db      536 CNLYDGEFREBNESISVCEDPQCEKXEDGLTCHGGPNCTKCSHFKGPPCVERCPD 595
Qy      603 GVPKDLSPYMPKPEDEGACOPCPINCTHSCVDLDKGC-----PAQCRASPL 651
Db      596 GLQGANF--IFXADRDRECHPHRCTGCGNCPHTSDCIYYPWTGSHSLPQAR-TPL 652
Qy      652 TSIVSANV-GILLVVVLGVVFGILLIKRQOKIRKRYMRLLOETELVEPLTPSGAMPNQA 710
Db      653 --IAAGVIGGLFIIIVGLTFAYVVRKRSIK-KKRALRRL-ETELVEPLTPSGTAPNQA 708
Qy      711 QMRILKEFNFTVSFVLRVPRVSAHLETYKGIWIDGKNVKIPVAKYLRNTPSKAN 770
Db      709 QRLTKE-----TELKRVKVLGSGAGCTYKGIWVEGETVAKIPAKILNETGPKAN 762
Qy      771 KEILDEAYVAVGSSPYVSRLLGLCLTSTVOLVTLMPYCLLDHVENRGRLSODLN 830
Db      763 VEFMDEALIMASMDPHLVRLLGCLSPITQVLTQMLPHSGCLLEVYEHKDNIGSQLLN 822
Qy      831 WCMQIAGMGVLEVRVLRDLARNVLSPPNHVKTDDGLARLDDIDTEYHAGGKY 890
Db      823 WCVQIAGMGVLEVRVLRDLARNVLSPPNHVKTDDGLARLDDIDTEYHAGGKY 882
Qy      891 PIKMALESILRRRFTQSDVWSYGVTVWELMTFGAKYDGIAPAREIPDLLEKGERLPP 950
Db      883 PIKMALESILRRRFTQSDVWSYGVTVWELMTFGAKYDGIAPAREIPDLLEKGERLPP 942
Qy      951 PICTIDYVMIMVCMMDLSECRPREFRVSEFSMADPPRFVYIQNED-LGASPLDST 1009
Db      943 PICTIDYVMIMVCMMDLSECRPREFRVSEFSMADPPRFVYIQNED-LGASPLDST 1002
Qy      1010 FYSLIEDDDMDGLVDAEYLVLPQGFCDPPAPAGAGVHHHRSSTSGGDLTLGL 1069
Db      1003 FPCNLDEEDLEMDMDEYLV-CAENIRPP-----ITSNARLDSNRS-----EIGH 1050
Qy      1070 EPSEEEAPRS-----PLAP-SEGAGSDVFDGLGMGA 1100
Db      1051 SPPEAYTPMSGNPFVYRDGFAAEQGVSEYRAPTSITIPAPVAGGATIEIFDSCNGT 1110
Qy      1101 AKGLQSLPTDPSPLQRYSEDPVPLPS-----ETDGVVAVLUTSPQPEYVYNDVNR 1153
Db      1111 LKRPVAVHVEDSSTORYSADPTVFAPERSPRGELDEGIMTWKRXKKEYINPYE--- 1167
Qy      1154 QPSPREGPLPAARPAGATLERAKTSPGNGVYKDVAFGAVENPEYLTPOGAAPOP 1213
Db      1168 -----ENPFVSR-----KMGDLQ-----ALNDPEYHNAAG----- 1194
Qy      1214 HPPPA-----FSPAFLUYVWDQDPPRGA- 1238
Db      1195 -PKADEYVNEPLYLNTFANTLGKAEYLNKNILNMPKAKKAFDNDYWNHSLPRSTL 1253
Qy      1239 -PSTFKGPT-----AENPEYL 1255
Db      1254 QHEDYLOEYSTKYFYKONGRIRPIVAENPEYL 1265

```

```

DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN      ERBB4 OR TYRO-2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_Taxid=10116;
RN      [1]
RP      SOURCE FROM N.A.
RC      TISSUE=Heart;
RX      MEDLINE=98221155; PubMed=953078;
RA      Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,
RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 275:10261-10269(1998).
RN      [2]
RP      SOURCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=9122560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system."
RL      Neuron 6:691-704(1991).
RN      [3]
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frommert P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulins and their putative receptors, ErbB2 and
RT      ErbB3, is induced during Wallerian degeneration."
RL      J. Neurosci. 17:1642-1659(1997).
CC      -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb.ch).
CC      -----
DR      EMBL: AF041838; AAD08899.1;
DR      EMBL: U52531; AAC53051.1;
DR      HSSP: P11362; 1FGK.
DR      InterPro: IPR000494; EGFR_L domain.
DR      InterPro: IPR000719; Euk_Pkinase.
DR      InterPro: IPR002174; Furin-like.
DR      InterPro: IPR001245; Tyr_Pkinase.
DR      InterPro: IPR004019; YLP_motif.
DR      Pfam: PF00757; Furin-like; 1.
DR      Pfam: PF00069; Pkinase; 1.
DR      Pfam: PF01030; Recep_L_domain; 2.
DR      Pfam: PF02757; YLP_2_domain; 2.
DR      PRINTS: PR00109; TYRKINASE.
DR      ProDom: PD000001; Euk_Pkinase; 1.

```

RESULT 7
 ERBB4_RAT
 ID ERBB4_RAT STANDARD; PRT; 1308 AA.
 AC Q62956; Q922N7;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

DB 1195 --PEKAEDEVNEPLVNTTNAIGNAEYMKNSLSYPEKAKAFDNDPDYHNSLPERST 1252
 QY 1239 --PPSTFKGTP-----NENPEYL 1255
 DB 1253 LQHPDYLQEVSTKYFYKONGRIRPIVANEPEYL 1285

RESULT 8
 XMRK_XIPMA STANDARD; PRT; 1167 AA.
 ID XMRK_XIPMA
 AC P13388;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TY.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=30015140; PubMed=2797166;
 RA Wilbrod J., Adam D., Malitschek B., Maueier W., Raulf F.,
 RA Telling A., Robertson S.M., Schartl M.,
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RT inducing Tu locus in Xiphophorus";
 RL Nature 341:415-421(1989).
 RN [2]
 RP REVISION TO 515.
 RA Schartl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X16891; CAA34770.2; -
 DR PIR: S06142; S06142.
 DR HSBP; P11362; IFGK.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002290; Ser_thr_Pkinase.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00069; Pkinase.1.
 DR Pfam: PF00757; Furin-like.1.
 DR Pfam: PF01030; Recept_L_domain.2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_Pkinase.1.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00220; S_TKC; 1.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642 KINASE.
 FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 666 1167 POTENTIAL.
 FT NP BIND 710 977 CYTOPLASMIC (POTENTIAL).
 FT BINDING 716 724 PROTEIN KINASE.
 FT ACT SITE 743 743 ATP (BY SIMILARITY).
 FT DISULFID 743 743 ATP (BY SIMILARITY).
 FT DISULFID 835 835 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 236 BY SIMILARITY.
 FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 479380749DC1D55A CRC64;

Query Match 38.1%; Score 2698.5; DB 1; Length 1167;
 Best Local Similarity 44.2%; Pred. No. 1.5e-132;
 Matches 564; Conservative 167; Mismatches 396; Indels 149; Gaps 31;

QY 4 AALCRMGILALLPGCAST---QVGTGDMKRLRPASPEHLDMLRLYQCGQVQGN 59
 DB 8 AALLQ--LLLVLSISRCSTDPDRVCQGTSNQMTM---LDNHVLMKKMYGCGNVLEN 62
 QY 60 LETVLPYVASLPIODIOEVQGYLIANOVROVLORLIVRGTOFEDNYALAVLDN 119
 DB 63 LETTYQENQDLSFLOISIOEVGYLIANNEVSTIPLVNLRLRGQNLTEGNTLLVMSN 122
 QY 120 GDPNNTPVTGASPGGLRELOLRSTELIKGGLVLIQRNPOLCYODTILMKDIFKRNQL 179
 DB 123 YQK-NPSSP--DYGVLKQLQSLNLTILISGVKVSHPFLCNVETIWMWDIVXTSNP 179
 QY 180 AALLPDTNRSRACHSPWCKRSRCMGESSEDCOSTIRVYVCGGC-ARKKGLPLPDCCH 238
 DB 180 TNNLIPHAFERQCKDHQCVNGSCWAPRGHCQKFKLLCEQCNRRRGKRPDCCNE 239
 QY 239 OCAAGCTGPKSHDCLACLFHNSGICELCPALVLYNTDFESMNPREGYFGASCAYTA 298
 DB 240 HCAAGCTGPRATDCLACRFNDGCKDPCPKYIYIVSHQVNVNPNKIFYGACAYVE 299
 QY 299 CFYNYLSTVGSCTLVCPAHNOEYTAEDTQCEKCKSPCAVCTGLGMQYIKANSKFLG 358
 DB 300 CFSNVYVE-GACVNSCSAGMLEVD-ENKRSCKDGVCPYCGIGISL-SYTIWVN 356
 QY 359 ITEL-EFACKKIFSLAFLPESFDGDPASNTAPQBPOLQVETLEBEITGLYISAMPD 417
 DB 357 STNINSFSCTININDIILNRISFEQDPHYKGTMDPRLMLTLTKETIGLYLVMMWPE 416
 QY 418 SLPLDLSVFONLOVIRGLIHNGAYS-LTLQIGISLWGLRSRLREIGSGALTIHNTHLCF 476

```

Db      417 NMISLVSQNLIELIRGRTTFSRGSFVAVQVRHLQWGLRLKSVSGNVILKXTLQRLY 476
Qy      477 VHTVPMDQLFRNPQALHTANRPEDECEGLACHOLCAGHCHMGPGPTCCVNCOSQFLR 536
Db      477 ANTIMRRLTFSESDSIEYDART-----ENQCNNECESDGCWGFPTMVCSCLVHVR 529
Qy      537 GQCEVBCRVUQLGLPREYVNAHCLPCHPCECPQNGSVTCGPEADQCACAHKDPFC 596
Db      530 GGRCVASCNLLQGEFREAVQDRCVQHCECLVGTDSLTCCGPGPANCSSAKAFQDPQC 589
Qy      597 VAPCSGVKPDLSVYPIKPFPEEGACQPCINCTHSCVDLDDKGCAPACRASPITSVS 656
Db      590 IPRCHHGLSDGDTL-IRKADKMGQCCQCHQNCCTQSGSGSLSCGKD-IVSHSLAVG 647
Qy      657 AVGGLLVVLGVVFGILLIKRQOKIRKTYMRLLOETELVEPLTPSGANPQAKMILK 716
Db      648 LVSGLLIVLIVALLIVLLRRRRIRK-RKRTRCLLOEKELVEPLTPSGQANQAFRLILK 706
Qy      717 --EFNNFTVSFWLRVPKVSASHLETVYKGIWIPDGENKIPVALIKTLRENTSPANKEL 774
Db      707 ETEFKKRV-----LGSAGFVYKGLMNPGENIRITVALIKTLRENTSPANKEL 758
Qy      775 DEAYVAVGVSPYVSRLLICLTSTVQLVTLQMPYGLLDHVRNKGRLSODLLNMQ 834
Db      759 DEAYVAVGVDPHVCRLLIGICLTSAVQLTQMPYGLLDVROHOERIQQMILLNMCVQ 818
Qy      835 IAKGMSYEDRVLRHDLAARNVLYKSNHVKITDFGLARLDDIDETEHYADGGKVPILK 894
Db      819 IAKGNVYEEHRLHARDLAARNVLYKSNHVKITDFGLARLDDIDETEHYADGGKVPILK 878
Qy      895 MALESILRRFTHSDVMSYGVTYWELMTFGAKRYDGIIPAREIPDLLEKGERLPQPICT 954
Db      879 MALESILQWLYTHQSDVMSYGVTYWELMTFGSKRPDGIIPAKEIASVLENGERLPQPICT 938
Qy      955 IDVYVIMYKCMWIDSECRPRELVEFSRMAPRQRFVIVQNEDELGASPLDSTFTRSL 1014
Db      939 IEVYVIMYKCMWIDSECRPRELVEFSRMAPRQRFVIVQNEDELGASPLDSTFTRSL 995
Qy      1015 LEDDMGDLVAEEYLVQGFPCDPAPAGAGVWHHRSSSTRSGGDLTLLEPSEE 1074
Db      996 LSSDD-DVDADELVLKYRI-----NRQS----- 1020
Qy      1075 EAPRPLAPSGAGSDVDGLMGAAKGLSLPHDPSLQRISEDPV-PLPSETDGY 1133
Db      1021 --EPCIPPTGH-----PVRENSITLIRNISPQVALKDKLDGH 1056
Qy      1134 VAPLTCSPQPEYVNPQVPRQP-----PSPRE-----GPLP-AARPAGATLERAKTUS 1180
Db      1057 -----EYVNPQSETSRSLSDIYNPVYEDLTDGMPVLSLSCQAEINPSPREYLN 1106
Qy      1181 PGKGVGVKQVAFGCAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDODPPERGAP 1240
Db      1107 TNQNSL--PLVSSSMDDPY--QAG-----YQAF-----LPQTAL 1141
Qy      1241 STFKGTPTAENPEYLG 1256
Db      1142 GNGMFLPAENLNYLG 1157

```

```

CX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90083234; PubMed=2687875;
RA      Kraus M.H., Isasing M., Miki T., Popescu N.C., Aaronson S.A.;
RT      "Isolation and characterization of ERBB3, a third member of the
RT      ERBB/epidermal growth factor receptor family: evidence for
RT      overexpression in a subset of human mammary tumors."
RL      Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=90311312; PubMed=2164210;
RA      Plioman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA      Todaro G.J., Shoyab M.,
RT      "Molecular cloning and expression of an additional epidermal growth
RT      factor receptor-related gene."
RL      Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN      [3]
RP      SEQUENCE FROM N.A. (SHORT FORM).
RC      TISSUE=Placenta;
RX      MEDLINE=93282822; PubMed=7685162;
RA      Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT      "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT      tyrosine kinase."
RL      Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC      -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC      (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC      SECRETED (SHORT FORM).
CC      -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC      SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC      -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC      -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC      SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC      AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC      PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC      -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      "This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch)."
CC      -----
DR      EMBL; M29366; AAA35790.1; -
DR      EMBL; M34309; AAA35979.1; -
DR      EMBL; S61953; AAB26935.1; -
DR      PIR; A36223; A36223.
DR      HSSP; P11362; 1PKK.
DR      Genew; HGNC:3431; ERBB3.
DR      MIM; 190151; -.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00069; pkinase; 1.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      ProDom; PD000001; Euk_pkinase; 1.
DR      SMART; SM00261; FU; 3.
DR      SMART; SM00219; TYKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR      PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR      transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM

```

KW	Alternative splicing.	POTENTIAL.
FT SIGNAL	1 19	RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT CHAIN	20 1342	EXTRACELLULAR (POTENTIAL).
FT DOMAIN	20 643	POTENTIAL.
FT TRANSMEM	644 664	CYTOPLASMIC (POTENTIAL).
FT DOMAIN	665 1342	PROTEIN KINASE.
FT NP BIND	709 966	ATP (BY SIMILARITY).
FT BINDING	742 742	ATP (BY SIMILARITY).
FT ACT SITE	834 834	BY SIMILARITY.
FT DISULFID	186 194	BY SIMILARITY.
FT DISULFID	190 202	BY SIMILARITY.
FT DISULFID	210 218	BY SIMILARITY.
FT DISULFID	214 226	BY SIMILARITY.
FT DISULFID	227 235	BY SIMILARITY.
FT DISULFID	231 243	BY SIMILARITY.
FT DISULFID	246 255	BY SIMILARITY.
FT DISULFID	259 286	BY SIMILARITY.
FT DISULFID	290 301	BY SIMILARITY.
FT DISULFID	305 320	BY SIMILARITY.
FT DISULFID	323 327	BY SIMILARITY.
FT DISULFID	500 509	BY SIMILARITY.
FT DISULFID	504 517	BY SIMILARITY.
FT DISULFID	520 529	BY SIMILARITY.
FT DISULFID	533 549	BY SIMILARITY.
FT DISULFID	552 565	BY SIMILARITY.
FT DISULFID	556 573	BY SIMILARITY.
FT DISULFID	576 585	BY SIMILARITY.
FT DISULFID	589 610	BY SIMILARITY.
FT DISULFID	613 621	BY SIMILARITY.
FT DISULFID	617 629	BY SIMILARITY.
FT CARBOHYD	126 126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	250 250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	353 353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	408 408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	414 414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	437 437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	469 469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	522 522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	566 566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD	616 616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC	141 183	ELISGVYIEKNDICLMDITIDMRDIVARDIAEIVVNDGR SC -> GQFPMVPSGLTPQPADVYLDDDPRLTLTSSASK VPTLTAAY (IN SHORT ISOFORM).
FT VARSPLIC	184 1342	MISSING (IN SHORT ISOFORM).
FT CONFLICT	560 560	E -> G (IN REF. 2).
FT CONFLICT	1064 1064	E -> G (IN REF. 2).
FT SEQUENCE	1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;	
Query Match	34.2%; Score 2343.5; DB 1; Length 1342;	
Best Local Similarity	39.7%; Pred. No. 2.9e-118;	
Matches	524; Conservative 195; Mismatches 459; Indels 143; Gaps 36;	
QY	10 GLLALLPFGAA--STOYCTGTMKRLIPASPEFHLMRLHYGCGQVVOGNIETLYLPT 67	
DB	11 GLTSLARGSEVSGNSQAVCPGTNLGLSVTGAENQYOTLYLKYRCEVWGMNIEIVLTGH 70	
QY	68 NASPFLQDIOEVQGVYIAHNOYVQLRLIRVGRQLPQEDVYALAVLNDGPIUNNTT 127	
DB	71 NADSLFQWIREVGYLVANNESTLPLPVLIRVGRQYVDGAFALFVW-----LYNNT 125	
QY	128 PVTGASPGELRELQRLSTELILKGVLIQRNPOLCYOTILMKDIFKHNQALATLLIDTN 187	
DB	126 ----NSSHALRQLTLQTLTEILSGVYIEKNDICLMDITIDMRDIVARDI--AEIVKD 178	
QY	188 RSRACHPCSPMKCSKRCWGESSEDCOSLTRVCAAGC-ARCKGSLPDDCCHEQCAAGCTG 246	
DB	179 NGRSCPHEVCKG-RCMGPGSEDCQTLTKTCAPQCNCHCFGSENMQCCCHDECAAGCAG 237	
QY	247 PKHSDCLALHFNHSGICELHCPALVTYNTDTFESMNPBEGRTFGASCYTACPYNYLST 306	
DB	238 PCDTDCFACRHFNDSGACVPRCPQLVYVNTKLTFLERPNHTKXYGYGVCVASCYHNFV-V 296	

QY	307 DVGSCITLVCPILHNOETIADGTCRCCKSPCARVCYGLAMQYIKANSKF--IGITELE- 363	
DB	297 DQNSCYRACPRDMEVD-KNGLKMCPCGGLCPKAEGRG-----SGSRQYVDSNDIG 350	
QY	364 FACCCKI FGSIAFLPESPDGPASNTAPQPEQVYFETLEITGYLYISAMPDLSLDLS 423	
DB	351 FVACTILGLDELTLGLNCDPWHKIPALDPEKIANFRIVREITGYLIQSWPWHNFS 410	
QY	424 VPCNLQVIGRIILHNAVS--TLQGLIGIMLGRISREISGIALIHHNTHLGFVTVPM 482	
DB	411 VPSNLTITIGRSILYNGFSLLIMKLNVTSLGRSLKEISAGRIYISANQOLYHNSLM 470	
QY	483 DQFRPHQALHTA-NREDECVGBLACHOLCARHGMCPGPTOCVNCSPLRQECV 541	
DB	471 TKVLGRTEERLDIKNRPDRDVAEGKVCDDLCSGCGWGPFGQCLSCRNSRGVCV 530	
QY	542 EECRVLOGLPREVYNAHCLCPHPEQOPONGSTGCPREDDQVACAHYKDPFCVACRP 601	
DB	531 THCNFNGEEREPHAECEFCSPCEQMEGTATCNSSGSDTCAQCAHFRDGHCVSSCP 590	
QY	602 SGVPRPLSVYPIKFPDEBACQPCPINCTHSCVDLDDKCCPAEQRA---SPLTSIVA 657	
DB	591 HGVLG--AKGPIKYKFDVQNECRPCHEMCTQCGKPELDCLGQTLVLIGKHTLTALT 648	
QY	658 VGLILLVYVLGVFGILLIKRQOKI-KYTMRLQETLVEPLTSGAMPQAQMRILK 716	
DB	649 IAG--LVVIFPMIGTFELWGRRIQNKAMRRYLGRGSEIPLDPS-EXANVYLARIFK 705	
QY	717 EFNPNFVPSFLRVPKYSASHL-ETVYKGIWIPGSEVVKIPVALKYLRNTPSKANIELD 775	
DB	706 ETE-----LRKLKVLGSGVFTGKWIIPGESIKIPVCIKVIEDSGRSFOAVTD 758	
QY	776 EAYVMAGVSPYVSRLLGICLTSTVQVLTQMLPFGCLDHPVENRRLSGQDLLNMCQI 835	
DB	759 HMLATSLDHAHIVRLIGCPGSSQLQVLYPLGLDLHVRQHGALGQLLLNMGVQI 818	
QY	836 AKMSALEDVRLVHRLAANVTVKSPPNKKITDREGLARLDDIEHYADGSKVPIKMM 895	
DB	819 AKMYYLLEEGWHRLAANVTLKSPQVQVADLPVADLLPPDDQLYSEAKTPIKMM 878	
QY	896 ALESILRRRTTHOSDVSXGVTVMEIMTGAKPYDGIIPAREIPDLEKGRILPOPICTI 955	
DB	879 ALESIHFGKTHQSDVMSYGVTVMEIMTGAEIYALRLAEVDDLEKGRILAQOICTI 938	
QY	956 DVTMIVKCKMIDSECRPRRELVSFSRMAQDPQFVYIQNEDLQRA---SPLDSTFYR 1012	
DB	939 DVTMIVKCKMIDENIRPTFKELANEFTRWARDPRRYLVYIKRES-GGILAPGPEPHGLTN 997	
QY	1013 SLLEDMDGDLVAAEYLVPOQGFPCDDAPAGAGVHNRHRSSTRSGGDLTLGLEP- 1071	
DB	998 KXLEVELEBEDLDLDEAEED-----NLATTTIGSALSLEVGTL 1038	
QY	1072 SEEBARSPPLAPSEAGSDVFDGLMGAKGLQSLPTHD-PSPLQRYSEDPVTPP- 1127	
DB	1039 NRPQGSQSLSPSSGY-MPMNQNLGESCQESAVSSSSBRCPVSLH-----PMPRG 1091	
QY	1128 ----SEIDGYVA-----PLTSGQPE-----YVNPQVVRQPSRRCGP- 1162	
DB	1092 LASESEGHGTGSEALQEKVMSKRSRSPRPGDSAYHSQHSLLTPVPLSPGCI 1151	
QY	1163 -----LPAARPAATLIERAKTLP-GRKNV-----KDVFAFGAVENFEYLTQ 1206	
DB	1152 EEDVNGVYMPDTHLKGTPSSREGTLSSVGLSVLTBEDEDE-----DEYETWNR 1203	
QY	1207 GGAAPQHPHPAPSPAFDMLTYWD-----QDPPEKCAPSTKGTPTANPEY 1254	
DB	1204 RHRSP-PHPRPSLSEILEYETWVDGSLASLGSTQSCPLHPVPMPTAGTTPDEY 1262	
QY	1255 L 1255	
DB	1263 M 1263	

RESULT 10
 ERB3_RAT STANDARD: PRT: 1339 AA.
 AC Q62799; Q62955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (c-erbB3).
 OS Rattus norvegicus (Rat).
 GN ERB3.
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 RX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 recombinant protein.";
 RL Gene 165:279-284(1995).
 RN [2]
 RP REVISIONS TO 85, 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=930624;
 RA Carroll S.L., Miller M.L., Fichner P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroligins and their putative receptors, ErbB2 and
 ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGLINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = AMP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE ERB RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: U29339; AAC28498.2; -;
 DR EMBL: U52530; AAC3050.1; -;
 DR HSSP: P11362; TRK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR PRODOM: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal.

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT CYTOPLASMIC 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235
 FT DISULFID 231 243
 FT DISULFID 246 255
 FT DISULFID 259 286
 FT DISULFID 290 301
 FT DISULFID 305 320
 FT DISULFID 323 327
 FT DISULFID 500 509
 FT DISULFID 504 517
 FT DISULFID 520 529
 FT DISULFID 533 549
 FT DISULFID 556 573
 FT DISULFID 576 595
 FT DISULFID 589 610
 FT DISULFID 613 621
 FT DISULFID 617 629
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT CONFLICT 1028 1028
 FT SEQUENCE 1339 AA; 147545 MW; 0A5F2402BBDFE CRC64;
 Query Match 33.1%; Score 2271.5; DB 1; Length 1339;
 Best Local Similarity 40.0%; Pred. No. 2,1e-114;
 Matches 517; Conservative 171; Mismatches 437; Indels 167; Gaps 37;
 QY 3 LALCWMGULLLPPGAA---STVCTGTDMLRLPASPEHTLMDLRLHYGCGVVOGN 59
 DB 7 LQVLC---FLSLARGSEMGNSQAVCPGTLGLSTGTADNOYOTLLYLKECEVVMGN 62
 QY 60 LELTYLPNTASISFLDDIIEVQGVYLIANOVQVBLRLRIYRGTLQFEDNYALVDN 119
 DB 63 LEIVLGNHNDISFLQWIREVVGXYLVANNEFSVLEPLNLRVVRGTVYDGKAFIVM-- 120
 QY 120 GDPINNTTPVTGASPGGLRELOLRSTLELLKGVLIQRPPOLCYOTILMKOIFHNOL 179
 DB 121 ---LNTNT---NSSALROLKFTQLTELLSGVYIEKDKICHMDITIDRVR-- 170
 QY 180 ALTLIDNRSRACHPCSPMCKSGRCGSESSDDQSLTRVACAGG-CARCKGLPTDCH 238
 DB 171 GAEIVVKNNGANCPCHVECKG-RCMGPDPDQILTKITICAPQCGRCFGPNQCCHD 229
 QY 233 QCAAGCTGPRHSCLCLAFNNSGITELHCPALVTNTTFESMPPEERYTFGASCVTA 298
 DB 230 ECAAGGSGGDDTDCFCRRFNDSGACVPCPELVNKLTFULEPPIHTKYGGCVAS 289
 QY 299 CPYNYLSTVGSCTVLCPLHNOEVALEDTORCEKSKPCARVCYGL--GMOYIKANSKF 356
 DB 290 CPNRFV-VQGTFCVRCAPFDKHEVD-KHGLKNCEPCGGLCPKACSGTSGSRYQTWDSN 347
 QY 357 IGTLELFAKCKKIFGSLAFLPESFDPGDSAPNAPLQBPOLQVFTLEETIGLYLSAMP 416

```

Db 348 ID-----GFVWCTKILNLDPLITGLVNDPWHKIPALDPEKXANFRTREITGYLNIQSWP 403
Qy 417 DSLPLSLVFNOLQVIRGRILHNGAYS-LITOGISIMLGIRSRREISGLALIHNTHC 475
Db 404 PHMNSVSVSNLTITIGRSILYNGFSLLIKNLVSLGRSKELISAGRVIISAOQLC 463
Qy 476 FVHTVPMDOLEFRPHQALLHTA-NRPEDECVGEGLAHQICARGHMGPGPTQVCNCSQF 534
Db 464 YHSLNMTLRLLRGPSEERLDIKYDRPLGECLEAGKVCDDPLCSGGCGMGPGQCLSCRY 523
Qy 535 LRQGEVBEGRVQLPBEVYNARHCLPCHPEQPNQSVTCGEPADQCVACAKYKDP 594
Db 524 SRGACVTHCNFQGEPRFVHAQCPSCHPECLPEMGSTCGSSGSDACARAHFRDP 583
Qy 595 FCVACPSGVKPDLSYMPIMKFPDEGACOPCPINTHSC--VDLDDKCPAQRASPLT 652
Db 584 HCNVSCPHGILG--ANGPIKYPDAGNECRPHENHTQCGNGLQDCLGQAEVLMKSH 641
Qy 653 SIYSAVVGLLVVGLVGVFGILIKRQOKR-KYTRRLLOETELVEPLTPSGAMFNQAO 711
Db 642 LVIAVTVG--LAVILMLGGSFLYWRGRIONRARRYLREGESEIEPLDPS-EXANKYL 698
Qy 712 MRLLKEFNFTVSFMRVPSASHL-ETYYKGIWIPDGENYKIPVAIVLRBNTPKAN 770
Db 699 ARIFKETE-----LRKLKVLGSVFGVYHKGIVPEBESIKIPVCIVIEDKSGROS 751
Qy 771 KEILDAVYVAGVGSPPYSRLGICLTSTVOLVTOQMPYGLLDHVENRGRLGSDLLN 830
Db 752 QAVTDHMLAVGSLDHAHIVRLGLCGSSSLQVLTQYLPGLSLDHVKQRHETLPGQLLN 811
Qy 831 KCMQIAKGSYLEDAVLRDIAARVLYKSPHVVITDPGLARLIDIDETVHADGCV 890
Db 812 MGVOIAKKNYILEHSMVRDLALRNVMKLSQCVAVADPLPDDCKLHSEAT 871
Qy 891 PIKMALESILRRRFTHSDVWSYGVVWELMTFGAKPYDGIPIAEIPLLEKGRLP 950
Db 872 PIKMALESIHFGKYTHQSDVWSYGVVWELMTFGAEPYAGRLAEIPLLEKGRLP 931
Qy 951 PICTIDVYIMKCMWIDECRRFELVSEFSRMAKDRQVRVIVONEGLGASPLDSTF 1010
Db 932 QCTIDVYVWVKCMWIDINIRPFKELNEFTMAKDRPRLVTKRAS-GEQTP--PAA 988
Qy 1011 YSLLEDDMDGLVDAAEYLVPOQGFCDPAPAGAMVHHRRSSSTRSGGDTLLGE 1070
Db 989 BSVLTTEL-----QELLEPEL-----DLDLLE 1014
Qy 1071 PSEE-----EAPRPLAPBSG-----AGSDVFGDGLMGAAK 1102
Db 1015 ABEGLATLSGALSFTGTLLRPGOSGLSPSSGYMNMOSLGEACLDAAVAGREQ 1074
Qy 1103 GLQSLFTHDPSLQRYSEDPVPLPSETGVY---APL-----TC-----SPOPE-- 1144
Db 1075 FERPLSLH-PIPRGR-----PSSSSGHTTGEAEIOENVSCRSRSRSPRGD 1126
Qy 1145 --YVQPDVRFQPPSPREGP-----LPAARPAATLERAKTLP--GRNGV-- 1187
Db 1127 SAYHSQRHSILTPVPLSPGLEEDNGVMPDTHLGASSRSRGTLSSVGLGTE 1186
Qy 1188 ---KQVFATGAVENPEVLTPOGAAPQHP 1216
Db 1187 EEDDED-----EYEVNKRKRQSP-PRPP 1209

```

RESULT 11
 EGFR_DROME STANDARD: PRT: 1426 AA.
 AC P04412: O61601: Q9W2G0: P81868:
 DT 13-AUG-1997 (rel. 05, Created)
 DT 15-DEC-1998 (rel. 37, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
 (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).

GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Eterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
 RX MEDLINE=94350209; PubMed=8070664;
 RA Clifford R., Schubach T.,
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
 RT that several genetically defined classes of alleles cluster in
 RT subdomains of the receptor protein."
 RL Genetics 137:531-550 (1994).
 [2]
 RP REVISIONS.
 RA Clifford R., Schubach T.,
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE=85124611; PubMed=2982499;
 RA Lyven E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.,
 RT "The Drosophila EGF receptor gene homolog: conservation of both
 RT hormone binding and kinase domains."
 RL Cell 40:559-607 (1985).
 [4]
 RN SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
 RP STRAIN=Oregon-R; TISSUE=Embryo;
 RX MEDLINE=87002474; PubMed=3093080;
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.,
 RT "Alternative 5' exons and tissue-specific expression of the
 RT Drosophila EGF receptor homolog transcripts."
 RL Cell 46:1091-1101 (1986).
 [5]
 RN SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
 RP ANALYSIS.
 RX MEDLINE=99102120; PubMed=9882502;
 RA Leeshkin A.M., Yu S.-Y., Katz J., Baker N.E.,
 RT "Several levels of EGF receptor signaling during photoreceptor
 RT specification in wild-type, Ellipse, and null mutant Drosophila."
 RL Dev. Biol. 205:129-144 (1999).
 [6]
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Burton R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Planinich C., Baldwin D.,
 RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Baasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brooksstein P., Brothier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Doonan K., Dup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Hoack J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kemison U.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Klug D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris T., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RL "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 RP [7]
 RP SEQUENCE OF 959-1078 FROM N.A.
 RC STRAIN=Daekwanyeong;
 RX MEDLINE=85137938; PubMed=2983232;
 RA Wadsworth S.C., Vincent W.S. III, Biodeau-Wentworth D.,
 RT "A *Drosophila* genomic sequence with homology to human epidermal
 RT growth factor receptor.";
 RL Nature 314:178-180(1995).
 RN [8]
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 RP ANALYSIS
 RX MEDLINE=20389942; PubMed=1936959;
 RA Raz E., Schejter E.D., Shilo B.Z.,
 RT "interallelic complementation among DER/11b alleles: implications for
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";
 RL Genetics 129:191-201(1991).
 RN [9]
 RP REVIEW
 RX MEDLINE=97248481; PubMed=9094709;
 RA Perrimon N., Perkins L.A.,
 RT "there must be 50 ways to rule the signal: the case of the *Drosophila*
 RT EGF receptor.";
 RL Cell 89:13-16(1997).
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOSEROSA
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THORACIC AND ABDOMINAL GANGLIA.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052754; AAC08536.1; JOINED.
 DR EMBL; AF052753; AAC08536.1; JOINED.
 DR EMBL; AF052754; AAC08535.1; JOINED.
 DR EMBL; AF052752; AAC08535.1; JOINED.
 DR EMBL; K03054; AAA51462.1; -
 DR EMBL; K03417; AAA51460.1; -

DR EMBL; K03416; AAA50965.1; -
 DR EMBL; K03418; AAA51461.1; -
 DR EMBL; AF109077; AAD26134.1; -
 DR EMBL; AF109078; AAD26132.1; -
 DR EMBL; AF109082; AAD26132.1; JOINED.
 DR EMBL; AF109078; AAD26133.1; -
 DR EMBL; AF109084; AAD26133.1; JOINED.
 DR EMBL; AF109079; AAD26130.1; -
 DR EMBL; AF109081; AAD26130.1; JOINED.
 DR EMBL; AF109079; AAD26131.1; -
 DR EMBL; AF109083; AAD26131.1; JOINED.
 DR EMBL; AF109080; AAD26135.1; -
 DR EMBL; AE003454; AAF46732.1; -
 DR EMBL; X02293; CAA26157.1; -
 DR EMBL; X78920; CAA55523.1; -
 DR EMBL; X78918; CAA55521.1; -
 DR EMBL; X78919; CAA55522.1; -
 DR PIR; A00640; GQFE.
 DR HSSP; P11362; 1FGK.
 DR FLYbase; FBgn0003731; Egfr.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_Pkinase.
 DR Pfam; PF00069; pkinase.1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR PRODOM; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 7.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;
 KW Developmental protein.
 KM DEVELOPMENTAL PROTEIN.
 FT SIGNAL 1 30 POTENTIAL.
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 869 889 POTENTIAL.
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 938 1198 PROTEIN KINASE.
 FT NP_BIND 944 952 ATP (BY SIMILARITY).
 FT BINDING 971 971 ATP (BY SIMILARITY).
 FT ACT_SITE 1063 1063 BY SIMILARITY.
 FT MOD_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 27.4%; Score 1879; DB 1; Length 1426;
 Best Local Similarity 32.0%; Pred. No. 2.4e-93;
 Matches 460; Conservative 183; Mismatches 436; Indels 358; Gaps 44;
 QY 24 QVCTGTDMKRLPASPETHLDMRLHYGGCGVQGNLELYLPT-NASLSFLQDIQEVGG 82
 DB 100 KICITKSRLLSPSSKHEHYRLDRYNTCTVDGNLKLTLWLNENLDLSFLDNIREVTG 159
 QY 83 YTLIAHQVROVPLQRLATVGTQF-----EDNYALAVLNDGDLNNTTGTGSPGL 137
 DB 160 YLLISHVDKRVKVPFKLQIRKTLFSLSVEEKALV-----TSKM 203
 QY 138 RELQRLSTLEILKGVLLIQRPOLCYQDTILMKDIFHKKNQALATLLIDNRSRACHPSP 197
 DB 204 YLLEIPDLRDVLNGVGHNNYLNCHMTIQSEIVSNGTDAYVYDTPAEREKPKCHE 263
 QY 199 MKSGRCNGESSEDDQSLIRYVAGGCA--RCKGLPLPDDCHEQCAACTGPKPSDCILAC 255
 DB 264 SCTHG-CWGEGRKNGQKFSKLTCSPOCGAGCYGKPRECHLFPAGGCTGPTQKDCIAC 322
 QY 256 LHFNSGICELCPALVTYNTDTPFSMPNDEGRYTFGASCYACPYNYLSDVGSCTIAC 315
 DB 323 KNFPDEAVSKSECPMRKYNPTTYLLEINPBGKAYAGATVCKECP-GHLLNDNAGCAVSC 381
 QY 316 PLHNEVTAEDGTQCEKCSKPCAEVVCYGLGMQYIKANSKFIGITEL-----EPAGCKK 369


```

Db 382 PDKRMKGGE-----CPCNGPCPKTC-----PGTVLHAAGNISFNCTV 422
Qy 370 IFGSLAFPSFDS--DPAVNTA-----PLQBPOLQVFTLEITGYLSPMPDLPD 421
Db 423 IDGNIRILDTFSGFQDVYANYTWPRIYFLDPERFVSTVEITGYLNEGTHTPOFEN 482
Qy 422 LSVQNLQVIRGLIHNGAV-SLTLOGLSIMGLRSLRSLGSLALIHNTLTCVHTV 480
Db 483 LSTRNLETHHGRGLMSFAALAIYKSSLYSLMRLLKQISSGVITQNRDLCTVSN 542
Qy 481 PWDGFNPHQALLHTANRPEDECVGGLACHOLCARGHGWGPPTQVYNSQFLRGQEC 540
Db 543 RMPAIQKEPEOKVWVWNTLRADLCERKNGTICSDQCNEDGCGAGTQCLTKAFNENGTG 602
Qy 541 VEEGRVLOGLPREVYVNRHCLPCHPECOQNGSVTCGPPADQCVACAHKDPFCVARC 600
Db 603 IADGVYSNAVK--FDNRKTCIKHPCR-----TCNGAGADHCQECVHVDQHCVCSEC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKHYNDRGVCRECHATCDGCTGPKDTIGAGCTTNALINNDATVRCGLLKDKCPD- 713
Qy 609 SYPIWKF--PDEEGACQ-----CPI-----NCTH-----632
Db 714 GY-FWEYVPHQEGSLKPLAGRAVCRKCHPLCELCTNYGHEOVCSKCTHYKRQOCT 771
Qy 633 -----SC-----VDLDKG-----641
Db 772 ECPADHYTDEQRCEQFHRHPCNGCTGPGADCKSCNFTLPANEGRPVNSMTNCTNS 831
Qy 642 -CPAEOR-----ASPLTS-----TSAVVGILLVVLGVVFGI 673
Db 832 KCPLEKHNHYOYATLGPYCAASEPPSSKITANDVMIRIIGAVLPTICILCV--T 889
Qy 674 LIRROQKIRKYT--MRLLQETELVPLTPSGAMPQAMRIKEKNNTVTSFWLRVPR 731
Db 890 YICRQKQKAKKETVMTALISGCEDESPLRPSNIGANLCKRLIVKDE-----LRKG 942
Qy 732 V-SASHLETYKGIWIDGDNVXIPIAIVKLYRENTSPKANKELIDEXYVAGVSPYVS 790
Db 943 VLMGARGRYKGVWVEGENVXIPVAKIKELKSTGASESEEFREXYVASEHVNLLK 1002
Qy 791 LIGICLTSTVQVLTQMPYGLLDHVENRGRIGSQDLMWCQJAKGMYLDEVLNHR 850
Db 1003 LLAVCSSQMLLITQLPLGLGLVYANNMDKIGSKALMLWSTQIAGKSYLEKRLVHR 1062
Qy 851 DLAARNVLKSPNHVKITDFGLARLIDIDETEHADGKVPKIMMALESILRRRFTHSD 910
Db 1063 DLAARNVLTOTPSLVKITDFGLAKLSSDSENYKAAGKMPKIMALECIRNRFVTSKD 1122
Qy 911 VMSYGVTVMLMTFGAIPYDGIIPAREIPDLLEKGERLPORPCTIDVYMTVMCKMIDSE 970
Db 1123 VMAFGVTIMELTFGQPHENIPAKDIPDLIEVGLKLEOPEISLIDYCLISCMHIDA 1182
Qy 971 CRPRFELVSEFSKMDPQRFVYIQNEDLG--PASPLDSTFYRSLIED--DMGLVD 1025
Db 1183 MRPFKQULTVFAEFARDPGRYALIPGDKXTRLP-----YTSQEKRLIKLAPTTD 1235
Qy 1026 AEEYLVPOQGFPCDPAPAGAGKVVHRHRSSTRSGGDLTLGLEPSEEFAP-----RS 1079
Db 1236 GSEAIAPDYLQPKAPPS-----HRTDCT-----BEMPKLNRXCAD 1274
Qy 1080 PLAPSEGAGSDVFDG--DLGMAKAGLQSLPTHDPSPFLQYSEDPVLPSPSTDPGVAP 1136
Db 1275 PSKNSSGTGDDERDSARREYGVNLR-----LDLPVEDDYLP 1313
Qy 1137 LTGSPQPEYVQDPVRQPSRPGPLPAARPAATLERAKTUSPGKNGVAVKVFAGGA 1196
Db 1314 -TCQPGNNNNNNNN--NPNQNNMAAVGAAGV-----DLIGVPS 1351
Qy 1197 VENPEYL-----TPQGAAPQPH-----PPAFSP-AEDNLYWD 1230

```

```

Db 1352 VDNPEYLNAQTLGVGSEPIPTQITGIPWGGPGTMEYKVPWPGSEPTSSDHEYYND 1408
RESULT 12
ID ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_Taxid=11864;
RN [1]
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman P.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated BGF receptor."
RL Cell 41:715-726 (1985).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIA ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: M10066; AAA48763.1; ALT_INTT.
DR PIR: A00643; TVCHLV.
DR PIR: B00643; TVCHLV.
DR HSSP: P13362; 1FGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR PRODOM: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TYRKc; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
KW Transferase; tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 24.3%; Score 1668.5; DB 1; Length 634;
Best Local Similarity 50.2%; Pred. No. 1.8e-82;
Matches 358; Conservative 83; Mismatches 143; Indels 129; Gaps 18;
Qy 587 CAHYKDPFCVAPRCPSGVKPLDSYPIWKFPEEGACQPCPINTCHSCVNDLDDKGPAPQ 646
Db 3 CAHFDGHCVKACACASVLTGENDTL-VKHYDANAANVQLCHPNCCTRGKSGELBSCP--- 58
Qy 647 RASPLTSLVSAVV-GILLVVLGVVFGILIRROQKIRKYTRRLLQETELVEPLTPSGA 705

```



```

Db      59  NGKSTPSIAAGVGGCLLVAVGIGLYLRRR-HIVKRTLRLLQRELVPLPSGE 117
Qy      706  MPNOQMRLIKEFNNFTVSFWLRPKVYASASHELVYKGIIPDGENVKIPVAIKVLRNT 765
Db      118  APNOAHRLIKE-----TEFKVKVVGSGAFGVYKGLIPEGEKVKIPVAIKELREAT 171
Qy      766  SPKANKEILDEAYMAVGSPPYVSRLLGLCTSTVQVLTQMLMPYGLLDHVRNRRGLS 825
Db      172  SPKANKEILDEAYMAVGSVDNPHVCRLLGLCTSTVQVLTQMLMPYGLLDHVRNRRGLS 231
Qy      826  QDLINMCQIAKGSYSLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 885
Db      232  QYLLNMCVQIAKGMVLYERRRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 291
Qy      886  DGGKVPFKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 945
Db      292  EGGKVPFKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 351
Qy      946  RLPOPICTIDVYIMTKCMWIDSECRPFRELVSEFSRMAPDPRFVVIQ-NEDLGPAS 1004
Db      352  RLPOPICTIDVYIMTKCMWIDSECRPFRELVSEFSRMAPDPRFVVIQ-NEDLGPAS 411
Qy      1005  PLDSTFYRSLIEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMVAHHRHRSSTSGGSD 1064
Db      412  PTDSEFYRSLIEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMVAHHRHRSSTSGGSD 449
Qy      1065  LTLLPSESEEPAPSPF-----APSEGAGSDVPFDGLDMGAKGLOSPTHDSPLOYS 1119
Db      450  -----SKTPLLSSLSATSNNSAINTCID-----RNGQSHPRFDSFYORIS 489
Qy      1120  EDPTVPLPSET--DGYVAPLTCSPQPEYVNOQPDVRCPPSPREGPLPAPAPAGATLERAK 1177
Db      490  SDPTGNLEESIDGFL-----PAPEYVNO--LMPKPKS-----521
Qy      1178  TLSFGKNGVYKDV-----AFGQVNEPYLTPQGAAPRHPAPASPA 1222
Db      522  -TAVVQNIYNNISLTAISKLPMDRYQNSHSTVDNPEYL-----NTNOSPILAKTV 572
Qy      1223  FDNLYWDC-----DPEE-----RGAPSTFGTPTAENPEYLGLDVP 1260
Db      573  FESSPWYIQSGNHOINIDNPYQODFLPNFKPKGLIKVPALENPEYLRLVAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
ID ERBB_AVIER
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain Es4).
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-76 (1983).
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=84221957; PubMed=6328658;
RA Deulite B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbb gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459 (1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC      tyrosine phosphate.
CC      -1- DISBASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC      ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC      IN CHICKENS.
CC      -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC      RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      EMBL; K02006; AAA42394.1; ALT_INIT.
CC      EMBL; K01216; AAA42400.1; -.
CC      DR PIR; A00644; TVYU.
CC      DR HSSP; P11362; 1FGK.
CC      DR InterPro; IPR000719; Euk_pkinase.
CC      DR InterPro; IPR001245; Tyr_pkinase.
CC      DR Pfam; PF00069; pkinase; 1.
CC      DR ProDom; PD000001; Euk_pkinase; 1.
CC      DR SMART; SM00219; Tyrc; 1.
CC      DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC      DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC      DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
CC      KW transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      glycoprotein; Phosphorylation.
CC      FT DOMAIN 132 399
CC      FT NP_BIND 138 146 ATP (BY SIMILARITY).
CC      FT BINDING 165 165 ATP (BY SIMILARITY).
CC      FT ACT_SITE 257 257 BY SIMILARITY.
CC      FT CONFLICT 29 29 S -> W (IN REF. 2).
CC      FT CONFLICT 140 140 S -> F (IN REF. 2).
CC      FT CONFLICT 146 146 I -> V (IN REF. 2).
CC      SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match: 23.7%; Score 1622; DB 1; Length 604;
Best local similarity 50.0%; Pred. No. 5; 3e-80;
Matches 348; Conservative 80; Mismatches 136; Indels 133; Gaps 17;

Qy      587  CAHYKDPPTCVARCSGKVPDLSTYPIKFPDEBAGCQCPNCTHSCVDLDDKCPAEQ 646
Db      3  CAHFIDGPHCVACAGVLENDTL-VKRYADANAVCQCPNCTRGCKGPGLECF-- 58
Qy      647  BASPLTSIVAVV-GILVVVGVGVGILIKERQOKIRKYTMRLLOETELVEPLTPSGA 705
Db      59  NGKSTPSIAAGVGGCLLVAVGIGLYLRRR-HIVKRTLRLLQRELVPLPSGE 117
Qy      706  MPNOQMRLIKEFNNFTVSFWLRPKVYASASHELVYKGIIPDGENVKIPVAIKVLRNT 765
Db      118  APNOAHRLIKE-----TEFKVKVVGSGAFGVYKGLIPEGEKVKIPVAIKELREAT 171
Qy      766  SPKANKEILDEAYMAVGSPPYVSRLLGLCTSTVQVLTQMLMPYGLLDHVRNRRGLS 825
Db      172  SPKANKEILDEAYMAVGSVDNPHVCRLLGLCTSTVQVLTQMLMPYGLLDHVRNRRGLS 231
Qy      826  QDLINMCQIAKGSYSLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 885
Db      232  QYLLNMCVQIAKGMVLYERRRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 291
Qy      886  DGGKVPFKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 945
Db      292  EGGKVPFKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 351
Qy      946  RLPOPICTIDVYIMTKCMWIDSECRPFRELVSEFSRMAPDPRFVVIQ-NEDLGPAS 1004
Db      352  RLPOPICTIDVYIMTKCMWIDSECRPFRELVSEFSRMAPDPRFVVIQ-NEDLGPAS 411
Qy      1005  PLDSTFYRSLIEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMVAHHRHRSSTSGGSD 1064

```

DB 412 PDSKFEYRLMEEDMEDIYDADEVLPVQGF-----NSPST----- 449
 QY 1065 LTGLPPESEARSP-----APSEAGSDVFDGDLGMAKGLSLPTHPSPLQRY 1119
 DB 450 -----SRTELSSLSTKSNNSATNCID-----RNGGHVREDSPYQKRS 489
 QY 1120 EDPVPLPSET--DGVAAPLTCSPPQPEYVNPDPVDPQPSFREGPLPAAPAGATLERAK 1177
 DB 490 SDPTGNFLEESIDGGL-----PAPEYVNO--LMPKKESTAM----- 524
 QY 1178 TLSPGKGVKGVAF-----CGAVENPEYLTPOGAPQPHPPA 1218
 DB 525 -----VQNTYNTSLTALSILKPLMSDRYQNSHSTAVDNPEYL-----NTNQSP 568
 QY 1219 FSPAFNLYVMDQPPERGAPSTFGKPTAENPEY 1254
 DB 569 AKTVFESSPYWIGSGNHQ-----INTLNDPXY 594

RESULT 14
 EGFR_CHICK STANDARD; PRT; 703 AA.
 AC P13387;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
 DE (fragment).
 GN EGFR.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88261272; PubMed=3260329;
 RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
 RA Ulrich A., Vennstrom B., Schlesinger J., Givol D.;
 RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
 RT transforming growth factor alpha.";
 RT Mol. Cell. Biol. 8:1970-1978(1988).
 RL Mol. Cell. Biol. 8:1970-1978(1988).
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M20386; AAA48760.1;
 CC InterPro: IPR000494; EGFR_L domain.
 CC InterPro: IPR000719; Euk_Pkinase.
 CC InterPro: IPR002174; Furin-like.
 CC InterPro: IPR001245; Tyr_Pkinase.
 CC Pfam: PF00757; Furin-like; 1.
 CC Pfam: PF01030; Recep_L domain; 2.
 CC SMART: SM00261; FU; 4.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; PARTIAL.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; PARTIAL.

DR PROSITE: PS0011; PROTEIN_KINASE_DOM; PARTIAL.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 30
 FT CHAIN 31 >703
 FT DOMAIN 31 654
 FT TRANSMEM 655 667
 FT DOMAIN 668 >703
 FT DISULFID 197 206
 FT DISULFID 201 214
 FT DISULFID 222 230
 FT DISULFID 226 238
 FT DISULFID 239 247
 FT DISULFID 243 255
 FT DISULFID 258 267
 FT DISULFID 271 298
 FT DISULFID 302 314
 FT DISULFID 318 333
 FT DISULFID 336 340
 FT DISULFID 513 522
 FT DISULFID 517 530
 FT DISULFID 533 542
 FT DISULFID 546 562
 FT DISULFID 565 581
 FT DISULFID 569 589
 FT DISULFID 592 601
 FT DISULFID 605 627
 FT DISULFID 630 638
 FT DISULFID 634 646
 FT CARBOHYD 134 134
 FT CARBOHYD 190 190
 FT CARBOHYD 200 200
 FT CARBOHYD 359 359
 FT CARBOHYD 359 359
 FT CARBOHYD 420 420
 FT CARBOHYD 573 573
 FT CARBOHYD 578 578
 FT CARBOHYD 613 613
 FT CARBOHYD 633 633
 FT CARBOHYD 648 648
 FT NON_TER 703
 SQ SEQUENCE 703 AA; 77427 MW; AFE2DE11B735A690 CRC64;

Query Match 23.3%; Score 1595; DB 1; Length 703;
 Best Local Similarity 44.6%; Pred. No. 1.8e-78;
 Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;

QY 8 RWGLLLALLPPGAA-----STOVCTGDMRLRUPASPEHLDMLRHLYCGCOVQGNLE 61
 DB 13 RGAATLVLLGLVALCSAVEEKVCOGNNKLTQLGHEHDFHTSLQRYNNCEVLSYLE 72
 QY 62 LTYLPYNASLFLDIOGVGVLIANQVQVPLQRIYRGTOLEFDNYALAVLQNGD 121
 DB 73 ITVEHNDLTFLEKTIQVAGVYIALNMVDVIPLENIQIIRGVLYXNSPALVLSNYH 132
 QY 122 PLNNTPTVPGASPGGLRELTSLTEIKGGVLTORNPOLCYODTILMKDIFHKNQAL 181
 DB 133 -MNTKQ-----GLRLPMKRLSEILINGVKISNPKKCMQMTVLNMDILDSRK-PL 182
 QY 182 TLID-TNBSRACHPCSPCKSRGWGESSDDCGSLRTVCAAGCA-RCKGPLPDDCCEQ 239
 DB 183 TVLDFASVLSSCPKHPNCTEDHGWGAGEQNCQTLTKVICAQCCSGRGGRVPSDCCHNQ 242
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDPESMNPESGRYFFGASCVTAC 299
 DB 243 CAAGCTGRESDDCLACRFDPDADCKTCPLVLYNPTTYMDVNPBCKYSFGATCVREC 302
 QY 300 PYNLTSDVGSCTIVCPLANQVTAEDTGRCEKSKRCARVCYGLQMYIKANSKFIGI 359
 DB 303 PHNYVVDHSGCVASCNNTDYEY--EENGVRCKKCDGLCSVNCNGIGIGELKGLIS-INA 360
 QY 360 TELE-FACCKRIIFSLAFIPESFDGDRASNTAPLQPELOVYFETLEETGLVLYISAMPDS 418

Db 361 TNIDSPKCTKINGVSLVLAFLGDAFTKTLPLDPKKLDVFRVKEISGFLLIQAMPDN 420

QY 419 LPLDSVFQNTQVIRGRLHNGAVSLTQGIISMLGRSRELGSGLALIHNTLCEVH 478

Db 421 ATDLVAFENLEIIIGRTKROGQSLAVNKLKISLGRSLKEISDGIAMKKNCLYAD 480

QY 479 TVPDDOLFRRPHQALLHTARPEDECVGEGLAQHQLCARQHCWGPPTQCVCNCSQFLRG 538

Db 481 TMRKSLFATOSQKTKIIONRNKQDCTADRHVCDPLCSDVGCWGPFGHFCSCRFPSRK 540

QY 539 ECVEECRVLOGLPREVYNAHCLPCHPECOPONG--SVTCFGEADQVACAHYKDPF 595

Db 541 ECVAQCNILQGEPEPEFRDSKCLPCHSECLVONSTAVNTTCSGPGPHCKMAHFDIGPH 600

QY 596 CVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDKCGCPARQASPLTSIV 655

Db 601 CVKACPGAVLGENDTL-VMKADANAVCQCHPCPTGCKGPGLEGCP--NSKTPSIA 656

QY 656 SAVV-GILLVVLGVVFGILIKRQOKIRKRYTMRLLQETELVEPLTP 702

Db 657 AGVVGGLLCLVVGVLGIGLYLRRR-HIVRKRTLRRLQERELVEPLTP 703

RESULT 15

ERBB_AVIEU STANDARD; PRT; 540 AA.

ID ERBB_AVIEU STANDARD; PRT; 540 AA.

AC P11273;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).

GN V-ERBB.

OS Avian erythroblastosis virus (strain ts167).

OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.

NCBI_TaxID=103898;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=87064458; PubMed=2878364;

RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;

RT "A single amino acid substitution in v-erbB confers a thermolabile phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells.";

RT Mol. Cell. Biol. 6:1751-1759(1986).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC -1- tyrosine phosphate.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL/M13179; AAA42401.1; -

DR PIR/A25231; TVFEVB.

DR HSSP/P11362; IRGK.

DR InterPro/IFR000719; Euk_kinase.

DR InterPro/IFR001245; Tyr_kinase.

DR Pfam/PF00069; Pkinase; 1.

DR ProDom/PD000001; Euk_pkinase; 1.

DR SMART/SM00219; TyKc; 1.

DR PROSITE/PS0107; PROTEIN KINASE ATP; 1.

DR PROSITE/PS0109; PROTEIN_KINASE_TYR; 1.

DR PROSITE/PS50011; PROTEIN_KINASE_DOM; 1.

DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;

KM Glycoprotein; Phosphorylation.

FT DOMAIN 132 399

FT NP_BIND 138 146

FT BINDING 165 165

FT ACT_SITE 257 257

FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).

SEQ SEQUENCE 540 AA; 60412 MW; 5853297AA068865D CRC64;

Query Match 22.6%; Score 1552; DB 1; Length 540;

Best Local Similarity 52.8%; Pred. No. 2.6e-76;

Matches 330; Conservative 72; Mismatches 127; Indels 96; Gaps 15;

QY 587 CAHYDPPPCVARSCEGVPDLSYMPIMKFPDEGACQPCPINCTHSCVDLDKCGCPAQ 646

Db 3 CAHFIDGPHCVACACAGVAGENDTL-VMKADANAVCQCHPCPTGCKGPGLEGCP 58

QY 647 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKRYTMRLLQETELVEPLTPSGA 705

Db 59 NSKTPSIAAGVVGGLLCLVVGVLGIGLYLRRR-HIVRKRTLRRLQERELVEPLTPSGE 117

QY 706 MENQGMRLKEFNNFTYSFMLRVPKVSASHLETYKGMIPDGENVKIPVAIKYLRNT 765

Db 118 APNQAHRLRLEK-----TEFKKRVLGEGAGFYKGMIPDEGEKVTIPVAIKYLRNT 171

QY 766 SPKANKELIDEAYVNAVGSPPVSRLLGICLTSTVQLVTQMLPVGCLLDHVAENGRGQS 825

Db 172 SPKANKELIDEAYVNAVGSPPVSRLLGICLTSTVQLVTQMLPVGCLLDHVAENGRGQS 231

QY 825 QDLNWCNOIANKGSLIEDVRLVHDDLAARNLVKSPNHYKITDGLARLDDIDETEHYA 885

Db 232 QYLNWCNOIANKGSLIEDVRLVHDDLAARNLVKSPNHYKITDGLARLDDIDETEHYA 291

QY 886 DGKVPYKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGE 945

Db 292 EGKVPYKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGE 351

QY 946 RLPQPPICITDYYMIMVKCMWIDSECRPFRELVSEFSRMAADPPQFVITQ-NEDLGPAAS 1004

Db 352 RLPQPPICITDYYMIMVKCMWIDSECRPFRELVSEFSRMAADPPQFVITQ-NEDLGPAAS 411

QY 1005 PLDSTFPRSLIEDDDMGDLVDAEYLVPPQGGFCDDPAPAGAGMWHRRRSSSTSGGSD 1064

Db 412 PLDSTFPRSLIEDDDMGDLVDAEYLVPPQGGFCDDPAPAGAGMWHRRRSSSTSGGSD 449

QY 1065 LTLGLEPSEEBARSPPL-----APSEGASDVFDDDLGMAKGLQSLPTHDPSPLOYYS 1119

Db 450 LTLGLEPSEEBARSPPL-----APSEGASDVFDDDLGMAKGLQSLPTHDPSPLOYYS 476

QY 1120 EDPTVPLPSETGTYAPLTCSQPEYVNAQPDVRRPPSPREGPLLAARAGAT-LEKAT 1178

Db 477 EDPTVPLPSETGTYAPLTCSQPEYVNAQPDVRRPPSPREGPLLAARAGAT-LEKAT 521

QY 1179 LSPGKGVVCDVFAFGAVENPEYL 1203

Db 522 DSRYN-----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:18:43

Job time : 22.6928 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 47.0485 Seconds

(Without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-695-709-14

Perfect score: 6853
Sequence: 1 MELALCRWGLLALPPGA.....TRKGPTEAPNPEYLGLDVPV 1261

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_ROTENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6130	89.4	1259	6	018735 canis fam1
2	3063	44.7	1209	11	090X70 ratu mus nov
3	3063	44.3	1210	11	090X70 ratu mus nov
4	2650	38.7	1165	13	09YH40 xiphophorus
5	2612.5	38.1	1137	13	09W6F6 gallus gall
6	2226	32.5	1328	13	09W6F6 gallus gall
7	1959.5	28.6	1433	5	09B1H9 fugu rubrip
8	1871	27.3	419	4	09UK79 mus sapien
9	1739	25.4	357	11	08R2X1 mus sapien
10	1697.5	24.8	412	4	08WYV0 mus sapien
11	1635	23.9	729	15	086712 avian rous-
12	1637	23.9	567	15	086714 avian rous-
13	1575.5	23.0	962	15	064895 avian eryth
14	1567	22.9	545	15	085468 avian eryth
15	1506.5	22.0	655	11	09WVFS mus sapien
16	1490.5	21.7	643	11	09ERV6 mus sapien

17	1204	17.6	1193	5	09Y1X8 ephydactia f
18	1139.5	16.6	1368	5	023821 caenorhabdi
19	1129	16.5	1717	5	026566 schistosoma
20	1126	16.4	527	13	090836 gallus gall
21	1001.5	14.6	478	11	09R5E0 ratu mus nov
22	942.5	13.8	599	13	09PSH2 gallus gall
23	906	13.2	165	4	014256 homo sapien
24	887	12.9	176	11	0923V5 ratu mus nov
25	806.5	11.8	346	13	091776 xiphophorus
26	778	11.4	435	5	08S2W1 drosophila
27	754.5	11.0	311	13	099162 xiphophorus
28	734	10.7	331	13	09BUD7 homo sapien
29	723	10.6	149	6	09B6G6 cycloctagus
30	705.5	10.3	1362	5	09PVZ4 xenopus lae
31	687	10.0	1671	13	09NJV5 biophalar
32	653.5	9.5	1368	13	08UM85 paratichthy
33	645	9.4	1418	13	093457 scophthalmu
34	640.5	9.3	1369	13	08UM86 paratichthy
35	626.5	9.2	1358	13	073798 xenopus lae
36	611	8.9	1472	5	09U5A8 bombux mori
37	601	8.8	1245	13	08UM84 paratichthy
38	597.5	8.7	1418	13	09YGH8 scophthalmu
39	587.5	8.6	2144	5	08UM83 paratichthy
40	577	8.4	1371	11	09QVW4 ratu mus nov
41	572.5	8.4	1091	4	09UMQ4 homo sapien
42	564	8.2	1072	4	091YMO mus sapien
43	562.5	8.2	1072	4	09BTB0 homo sapien
44	561	8.2	935	4	09EL35 mus sapien
45					

ALIGNMENTS

RESULT 1
ID 018735 PRELIMINARY; PRT: 1259 AA.
AC 018735;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erdb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erdb-2 from canine mammary gland";
RL Submitted (0CT-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB008451; EMBL2127.1; --
DR HSSP; P11362; 1FCR.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR L domain.
DR InterPro; IPR000719; Euk Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2_domain; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SMO0261; FU; 3.
DR SMART; SMO0219; TyrKc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase
SQ SEQUENCE 1259 AA; 137989 MW; E3764D9CA4CD46 CRC64;

Query Match 89.4%; Score 6130; DB 6; Length 1259;
 Best Local Similarity 89.6%; Pred. No. 0;
 Matches 1135; Conservative 43; Mismatches 75; Indels 14; Gaps 4;

QY 1 MELALCRWGLLLALPPGAASIVCTGTDMKLELPASPTHTDMLHLYOGQVVOGNTL 60
 DB 1 MELAMCRWGLLLALPPGAAGTCTGTDMKLELPASPTHTDMLHLYOGQVVOGNTL 60
 QY 61 ELTYLPTNASTLFLQDIQEVGYVLIAMNOVROVPLRLIRIVGTQLFEDNYALAYLDNG 120
 DB 61 ELTYLPTNASTLFLQDIQEVGYVLIAMNOVROVPLRLIRIVGTQLFEDNYALAYLDNG 120
 QY 121 DPLNTTPTVGASGSGRLREQLRSITILKGVILIQNPOLCYQDTILMKDIFHKONQA 180
 DB 121 DPLNTTPTVGASGSGRLREQLRSITILKGVILIQNPOLCYQDTILMKDIFHKONQA 180
 QY 181 LTLIDTNRSAACHSCSMCKGSCRMGSESDQSLRTTVCAGGARCKGPLPTDCCHEQC 240
 DB 181 LTLIDTNRSAACHSCSMCKGSCRMGSESDQSLRTTVCAGGARCKGPLPTDCCHEQC 240
 QY 241 AAGCTGPKHSDCLALCFHNSGICELHCPALVYINTDTFESMPNPEGRTYFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLALCFHNSGICELHCPALVYINTDTFESMPNPEGRTYFGASCVTACP 300
 QY 301 YNYLSTDVSGCTVLCPLANOVTAEDETCCKSKRCARVCYGLMGIYKANSKFTGIT 360
 DB 301 YNYLSTDVSGCTVLCPLANOVTAEDETCCKSKRCARVCYGLMGIYKANSKFTGIT 360
 QY 361 ELRFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQVETELIETGYLYISAMPDSL 420
 DB 361 ELRFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQVETELIETGYLYISAMPDSL 420
 QY 421 DLSTFQNLQYIRGILNGAYSLTLOGLISMLGRSLRSLGSLALIHNTLHCVHTY 480
 DB 421 DLSTFQNLQYIRGILNGAYSLTLOGLISMLGRSLRSLGSLALIHNTLHCVHTY 480
 QY 481 PMDLPFNPHQALHTANREDECEVSGELACHQLCAGHCWGPPTCCVNSQGLRGOEC 540
 DB 481 PMDLPFNPHQALHTANREDECEVSGELACHQLCAGHCWGPPTCCVNSQGLRGOEC 540
 QY 541 VEECRVLOGLPREVYVNRHCLPCHPECPONGSYTCGPADOCVCAHYKDPFCVARC 600
 DB 541 VEECRVLOGLPREVYVNRHCLPCHPECPONGSYTCGPADOCVCAHYKDPFCVARC 600
 QY 601 PSQVGPDLSTWPIKFPDEEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660
 DB 601 PSQVGPDLSTWPIKFPDEEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660
 QY 660 PSQVGPDLSTWPIKFPDEEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660
 DB 660 PSQVGPDLSTWPIKFPDEEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660
 QY 721 FTYSFWLRVPKVSAS-HLETVYKGIWIPDGENYKIPALIVYLENTPSKANKILLDEAY 779
 DB 721 FTYSFWLRVPKVSAS-HLETVYKGIWIPDGENYKIPALIVYLENTPSKANKILLDEAY 779
 QY 779 FTYSFWLRVPKVSAS-HLETVYKGIWIPDGENYKIPALIVYLENTPSKANKILLDEAY 779
 DB 779 FTYSFWLRVPKVSAS-HLETVYKGIWIPDGENYKIPALIVYLENTPSKANKILLDEAY 779
 QY 840 SYLEDVRLVHRDLAARVLYKSPNHVYITDFGLARLIDITETVHADGGVPIKMAALS 899
 DB 840 SYLEDVRLVHRDLAARVLYKSPNHVYITDFGLARLIDITETVHADGGVPIKMAALS 899
 QY 899 SYLEDVRLVHRDLAARVLYKSPNHVYITDFGLARLIDITETVHADGGVPIKMAALS 899
 DB 899 SYLEDVRLVHRDLAARVLYKSPNHVYITDFGLARLIDITETVHADGGVPIKMAALS 899
 QY 900 ILRRPFTHSDVMSYGVTVWELMTFGAKPYDGIIPABEIPDLTKGERLPPPTCTIDVYM 959
 DB 900 ILRRPFTHSDVMSYGVTVWELMTFGAKPYDGIIPABEIPDLTKGERLPPPTCTIDVYM 959
 QY 959 ILRRPFTHSDVMSYGVTVWELMTFGAKPYDGIIPABEIPDLTKGERLPPPTCTIDVYM 959
 DB 959 ILRRPFTHSDVMSYGVTVWELMTFGAKPYDGIIPABEIPDLTKGERLPPPTCTIDVYM 959
 QY 960 IMKCKMIDSECRPRFELVSEFSRMAKDPQRVYVIONELGASPLDSTFTFSLLEDD 1019
 DB 960 IMKCKMIDSECRPRFELVSEFSRMAKDPQRVYVIONELGASPLDSTFTFSLLEDD 1019
 QY 1019 IMKCKMIDSECRPRFELVSEFSRMAKDPQRVYVIONELGASPLDSTFTFSLLEDD 1019
 DB 1019 IMKCKMIDSECRPRFELVSEFSRMAKDPQRVYVIONELGASPLDSTFTFSLLEDD 1019
 QY 1020 MGLVDAAEYLVPOQGFCCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEBAPRS 1079
 DB 1020 MGLVDAAEYLVPOQGFCCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEBAPRS 1079

DB 1013 MODVDAAEYLVPOQGFCCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEBAPRS 1072
 QY 1080 PLASEBAGSDVDFDDLMGAAGKIQSLPTHDPSELQYSEDPYPLPSETDGYAPLTC 1139
 DB 1073 PLASEBAGSDVDFDDLMGAAGKIQSLPTHDPSELQYSEDPYPLPSETDGYAPLTC 1132
 QY 1140 SPOPEYVNOVDVPPSPREGLPLPAAPAGATLER-----AKTLPKNGVVKDVFAFG 1194
 DB 1133 SPOPEYVNOVDVPPSPREGLPLPAAPAGATLER-----AKTLPKNGVVKDVFAFG 1192
 QY 1195 GAVENPEYLTPOGGAAPQHPPPAPFADNLVYWDOPPERGAPSTFKTPTAENPEY 1254
 DB 1193 SAVENPEYLTPOGGAAPQHPPPAPFADNLVYWDOPPERGAPSTFKTPTAENPEY 1252
 QY 1255 LGLDVAV 1261
 DB 1253 LGLDVAV 1259

RESULT 2
 ID 090X70 PRELIMINARY; PRT: 1209 AA.
 AC 090X70;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=900258688; PubMed=2342466;
 RA Petch L.A., Harris J., Raymond V.W., Blaabard A.J., Lee D.C.,
 RA Earp H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue.";
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K., Dawson T.L., Earp H.S.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: M37394; AAF:4008.1; -.
 DR HSSP: P11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk pkinase.
 DR InterPro: IPR002174; Furlin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PR00757; Furlin-like; 1.
 DR Pfam: PR00069; Pkinase; 1.
 DR Pfam: PR01030; Recep_L domain; 2.
 DR PRINTS: PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Receptor; Tyrosine-kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE37F6CC1B773 CRC64;

Query Match 44.7%; Score 3063; DB 11; Length 1209;
 Best Local Similarity 49.1%; Pred. No. 1.5e-22;

Matches: 631; Conservative 169; Mismatches 362; Indels 122; Gaps 29;

QY 3 LALACRGLALLPEGA--ASTQVCTGTDMKLTLPASPETHLMDLHYOGCVGAGNLE 61
 Db 15 LALLCAAG-----GALEBKVCQGTNRRLTQGLTFEDHFLSLORMNCEVULGNIE 66
 QY 62 LITLPTNAGSLFQDIQEOGVYLIANNOVROVFLORLIVRGTQLFEDNYALAVLNDG 121
 Db 67 ITVQRNYDLSFKTLTQEVAGVLIANTVERIPELNQIIRNNALEYNTVALAVLSN-- 124
 QY 122 PLNNTPVTGASDGLRELQRLSTELKGLVLIQRNPOCYQDITLMDIFHKNQJLAL 181
 Db 125 -----YGTNKTGLBELPMRNDQELLIGAVRFSNNPLICMNETIQMDIV QDVFVLSN 175
 QY 182 TLIDITNRS-RACHPCSPMCGSRGWESSEDCQSLTRVYACAGCA-RCKGRLPTDCHEQ 239
 Db 176 MSMDVORHLTGCPKCDPSCPNGSCMGREBNCOKLTIKICACQCSRCRGSRPSDCCHNQ 235
 QY 240 CAGCTGPKRSDCLACHFNHSGICELHCPALVTYNTDTPESMPNREGRYTGASCVTAC 299
 Db 236 CAGCTGPRESDCLVCHRFDEATCQDTCPLMLYPTTYQMDVNEGKYSFGATCVKXC 295
 QY 300 PNYLSTDVGSCTLVPLHNOEYTAEDGTQRCCKSKPCARVCYGLGMOYIRANSKFIGI 359
 Db 296 PRYVYVTDHSGCVRACGPDYEV-EEDGVSCKCKDQCPKRYCNNGIGIEPK-DILSTIA 353
 QY 360 TELE-FAGCKKITGSLAFLESPDGPASTAPLOEQLOVETLEITGYLYISMPDS 418
 Db 354 TIKHFKYCAISGDIHLIPVAFKGSFRTPLDRELEIKTVEITIGFLLIQAMPEN 413
 QY 419 LPDLVFOQLQVIRGRIHLNGAYSLTLOGIGIMLGRSLRELGSALTIHNTHLCFVH 478
 Db 414 WDLHAFFENELIRGRTKOHGQSLAVGNTSLRLSKELISDDVILISGRNLCYAN 473
 QY 479 TVPMDOLFNRPHOALHTANRPDECVSGGLACHQICARHCWGPPTQCVNCSQFLRQ 538
 Db 474 TIKMKLFGTPNOKTIMNRRAEKDKATNHYCNPLCSSEGCSEGPPTQCVSQNYSRGR 533
 QY 539 ECVEEGRVIGLREYVNAHCLPCHPEGQPMGSGTCGFPADQVCAAHKDPFVCA 598
 Db 534 ECVDKNILEGEBREVENSECTQCHPECTPOTMNTCTGRGDNCTIKCAHVVDGHCVK 593
 QY 599 RCPGSKPDLSTYMPWKFPDEGACQPCPINCTHSCVDLDDKCPAQRASP-LTISVA 657
 Db 594 TCPSGIMGERNTL-VKFEADANNVCHLCHANCTYGCAGPLKCC-QDEGPKRIPSIANG 650
 QY 658 VNGIILVVLGVVFGI-LIKRQOKIRKXTMRLLQETLVELTSSGAMPNQAOQRIK 716
 Db 651 IVGGLLPTIV-VALGIGLEFMRRLQVRKTLRLLOERLEVELTSSGAPNQAHLRIK 709
 QY 717 EFNNFTVSEMLRVPKYSASHLETIVYKIMIPDGENYKIPVALIKVIRENTSPKANKELDE 776
 Db 710 E-----TEFKKIKVIGSAGFVYKGLMIPBEGEKKIPVALIKELREALSPANKELIDE 763
 QY 777 AYMAVAGSPYVSRLLIGITCLTSTVQVLTQMPYGLLDHYRENRGRSLQDILLNMCQIA 836
 Db 764 AYMAVAVDPHVCORLLIGITCLTSTVQVLTQMPYGLLDHYRENRGRSLQDILLNMCQIA 823
 QY 837 KGMSTYEDVRLVRLDAARVLAARVLAARVLAARVLAARVLAARVLAARVLAARVLAARV 896
 Db 824 KGMNVLIEDRLVRLDAARVLAARVLAARVLAARVLAARVLAARVLAARVLAARVLAARV 883
 QY 897 LESILRRFTHOSDVSAGVYVWELMTFGAKPYDGIIPAREIDLEKGERLPOPICTID 956
 Db 884 LESILHRIYTHOSDVSAGVYVWELMTFGSKPYDGIIPASEIISILEKGRRLPOPICTID 943
 QY 957 VYMIWYKCMWIDSECPREFRELVESESRADPQRFVVO-NEDGAPAPLUSTEYRSL 1015
 Db 944 VYMIWYKCMWIDSECPREFRELVESESRADPQRFVVO-NEDGAPAPLUSTEYRSL 1003
 QY 1016 EDDMGDLVDAEYVLPQGGFFCPDPAAGAGGVHHRHSSSTRSGGDLTLGLPSEEE 1075
 Db 1004 EEDMEDVVDADAEYLIPQGGFF-----NSPST----- 1030

QY 1076 APRSLAPSEGAGSDVFDGDLGMAKXQSLPTHPSPLORYSEDPVLPSEF--DGY 1133
 Db 1031 -SRTPLLSLSANSV-----SSTVACINNGSCRVKEDAFILQRYSSDPISVLTEDNIDTF 1085
 QY 1134 VAPLTCSQPEYVNOVDVRFPPSPRESQPLPAAPACGTLERAKTLSPGKGVYVDVAF 1193
 Db 1086 L-----VPEPINO-SVKKRPASGVONFVYHNOPLHP-----AGRDLHYON--PH 1128
 QY 1194 GGAIVENPEYL-TPQGAAPQHPAPFAPFADNLYVMDQ-----DP-----PER 1236
 Db 1129 SNAVENPEYLTAQ-----PCLSSGDSALWQKSHQMSLDNPDVYQDFFKE 1179
 QY 1237 GAPPSTFKGTPTAENPEYLGDPV 1260
 Db 1180 AKPNGIFKG-PTAEVAYELRVAVP 1202

RESULT 3
 Q9EP98
 ID Q9EP98 PRELIMINARY; PRT, 1210 AA.
 AC Q9EP98;
 DT 01-MAR-2001 (Tremblrel, 16, Created)
 DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
 DE Epidermal growth factor receptor isoform 1.
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Mable N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schenl C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Mable N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275366; AAC28045.1; JOINED.
 DR EMBL; AF275364; AAC28045.1; JOINED.
 DR EMBL; AF275365; AAC28045.1; JOINED.
 DR EMBL; AF275367; AAC24386.1; -
 DR HSSP; P11362; 1FGX
 DR MGD; MGI:95294; Egr.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; EGFR_L_domain.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TyrKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SMO0261; FU; 5.
 DR SMART; SMO0220; S_TKC; 1.
 DR SMART; SMO0219; TYKX; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 2.
 DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
 KW ATP-binding; Kinase; transferase; tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7F7EE38D771A74E CRC64;

Query Match 38.7%; Score 2650; DB 13; Length 1165;
 Best Local Similarity 44.7%; Pred. No. 3.3e-192;
 Matches 574; Conservative 163; Mismatches 389; Indels 158; Gaps 33;

QY 1 MELAALCRWGLLALLPPG-AAST-----QCTGDMALRLPASETHLDLRLHYQCCQY 55
 DB 4 LLELLEL-----LLELLLSIGRCSTDPDRKVCQGTSGNMTM---LDNHYLLMKKKYSGCNV 56
 QY 56 VQGNLELTYLPTNALSFLQDICEVGVYLIANQVAVPDLRLRYRGTLQLEEDNALA 115
 DB 57 VLENLEITYQENDLSFLQICEVGVYLIANQVAVPDLRLRYRGTLQLEEDNALA 116
 QY 116 VLDNGDPLNNTPYTGASPGGLRELQLRLSTELLKGVLLIQRPQLCYQDTLLMKDIFHK 175
 DB 117 VMSNYOK-NPSSP--DYVQVGLKQLOLSNLTLEISGVKVSHPNELLGNVETIMMDIVDX 173
 QY 176 NNQALTLIDTNRBACHPCSPCKSGSCWSESESDQSLTRTVCAAGC-ARCKGPLEPTD 234
 DB 174 TSNPTMILIPAFERQCKDPCGCVNGSCWAPRGHCKFTKLCAQCCKRRCKGPRPID 233
 QY 235 CCEBOCAAGCTGPKHSDCLAFHNSGICELHPCALVTYNTDTFESMPNPEGRTFGAS 294
 DB 234 CCEHCAGAGCTGPRATDCLACRDENDGTCCKTCCPKKIDIVSHQVANDPNIKYTFGAA 293
 QY 295 CVTACPNYLTSDVGSCTVCPHNGVTAEDGQREKCSKPARCYGLG-----QYI 350
 DB 294 CVKECFSNVYVTE-GACVRSQSGAMLEVD-ENGRSKRCPGCVGPKCDGIGISLSNTI 351
 QY 351 KANSKFIQILELPEACCKIFGSLAFLPESFDCPASNTAPLOEQVLEETLEITGYL 410
 DB 352 AVNSTNIG-----SESNCTKINGDIILNRNSPBGDPHYKIGMDEHMLNLTVEITGYL 407
 QY 411 YISAWPDSLPDLSFQNLQVIRGLINNGAYS-LTLOGLSWGLSLSLELGSGLALIH 469
 DB 408 VIMWPMENMTSLVFQNLLEIRGRTTSRGFSFVYGVSHLQWGLSLSLEVSAGNVLK 467
 QY 470 HNTLCEVHTVPMQQLFRNPHQALHTANREDECVESGLACHOLCARGCWGPGPTQCV 529
 DB 468 NTPQLRYASTINMRLEFRSEDSQILEYDART-----ENQCNNECEDDCWKGPRIMCV 520
 QY 530 NCSQFLRGQSCVEECRYLQGLPREYVNAHRLCPHPCQPPNGSVTCFEPADQCVACAH 589
 DB 521 SCHVDRGRCVASCNLLQGEPRBAOVDRGCVOCHOELVQTSILTCYGGPANCSCAH 580
 QY 590 YKDPFCVACRPSGVKPDLSYMPIMKFPDEEGACQPCPINTGSCVDLDDXGCAEQRAS 649
 DB 581 FQDPQCIIFRCPHMLDGDPTL-IMKADKMGQCPQCHQNTQSCSGPGLSGCRGD-IVS 638
 QY 650 PLTISAVAVGILLVVLGVVGLIKRRQCKIRKYMRLLOETLEVEPLTPSGAMPNQ 709
 DB 639 HSSLAIVGLVSGILITVIALIIVLLRRRIK-RKRTIRRLQCEKLEVEPLTPSGQAPNQ 697
 QY 710 AQMRILK--EFNNNTVSWFLVVKPVASASHLETYVKGIMWIPGEVVKIPVALIKYLRNTSP 767
 DB 698 AFLRLKEETFKKQRV-----LGSAGFVIVKGLMNPGENKIRIPVALIKYLRNTSP 749
 QY 768 KANKEILDEAVYVAVGSPYVSRLLGICLTSTVOLVTLQMPYGCILDHVENRGLSQD 827
 DB 750 KVNQGEVLDEAVYVAVSVHPVCRLLGICLTSAVOLVTLQMPYGCILDHVENRGLSQD 809
 QY 828 LLNWCQIAKGMSTLEDRVLRHRLAARNTVKSPPNVKITDFELALLLDIDEHYADG 887
 DB 810 LLNWCQIAKGMSTLEDRVLRHRLAARNTVKSPPNVKITDFELALLLDIDEHYADG 869
 QY 888 GKVPYIKMALESIRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPREIPLDLEKERT 947
 DB 870 GKVPYIKMALESIRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPREIPLDLEKERT 929

QY 948 POPPCTIDVYIMIMKCMWIDSECRPRFRELVSSEFSRMARDPQREVIQNEEDLGSPASPLD 1007
 DB 930 POPPCTIDVYIMIMKCMWIDSECRPRFRELVSSEFSRMARDPQREVIQNEEDLGSPASPLD 986
 QY 1008 STFYNSLLEDDMDGLVDAEYLVLPQQGFCFDDPAPAGAGVYHHRSSSTSGGDLTL 1067
 DB 987 RLFSRLSSSD--DVVDADDEVTL-----RYKRLN-RQGS----- 1018
 QY 1068 GLEPSEEARSPPLAPSEGAGSDVFDGLMGAAKGLSLPTNDSPLQRYSEDPV-PL 1126
 DB 1019 -----EPCIPPNCH-----PRENSIARLYSDPQNL 1047
 QY 1127 PSETDGYVAPLTCSPQPEYVNOPDYRPOP-----PSPRE-----GPLP-AARPAATL 1173
 DB 1048 EKDLDDH-----EYVNGPSEFTSSRLSDIYPNYEDLTDGMPVLSLSQEAETNF 1097
 QY 1174 ERAKTLSPKNGVAVDVFAGAVENPELTVQGGAAAPQPHPPAFSPPFDLVYWDQRP 1233
 DB 1098 SREYLNNTQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----L 1132
 QY 1234 PERGAPSTFKGTPTAENPEYLGL 1257
 DB 1133 PQTGALTGNGMFLPAENLEYGL 1156

RESULT 5
 Q9M6F6 PRELIMINARY; PRT; 1137 AA.
 ID Q9M6F6;
 AC Q9M6F6;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Receptor tyrosine kinase (Fragment).
 GN BRB94.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 ...
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A.,
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
 RT embryonic chick hindbrain";
 RL Mol. Cell. Neurosci. 13:237-258(1999).
 DR EMBL; AF121963; AAD31764.1; -.
 DR HSSP; P11362; 1FGK.
 DR Interpro; IPR000494; EGFR_L_domain.
 DR Interpro; IPR000719; Euk_pkinase.
 DR Interpro; IPR002174; Furin-like.
 DR Interpro; IPR001368; TNFR_c6.
 DR Interpro; IPR001245; Tyr_pkinase.
 DR Interpro; IPR004019; YLP_motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 1.
 DR Pfam; PF02757; YLP_2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00219; TyKc; 1.
 DR SMART; SM00261; FU; 3.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
 DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match 38.1%; Score 2612.5; DB 13; Length 1137;

Best Local Similarity 45.6%; Pred. No. 2.2e-189; Matches 523; Conservative 175; Mismatches 363; Indels 85; Gaps 27;

```

QY 161 LCYODITLWMDIFKNNQALATLIDTNRSPACHSPCKSGRCWGSSEDCOSLRTTVC 220
Db 3 LCFADTILHMDIVANPMASNFTLVPTNGSGCGRCHXCTG-FCMGFTENHCQLTFTVC 61
QY 221 AGGC-ARCKGSLPFDCCHEGCAACGTGPKSDCLACHPHNSGICELHCPALTYNTDTF 279
Db 62 AEGCDGRGYPVSDCCRCRCAGGSGCKDTCACMKFNDSGCVTQCQTFYVNTTF 121
QY 280 ESMNPBGRYTFGASCVTACPYNLSTDVSCCLVCPJHNOEYATDGTORCEKSKPCA 339
Db 122 QLEHNNAKYTYGAFVCKKCPHFV-VDSSCVACAPSSKREV-EENGIMCKCKCTDIPC 179
QY 340 RVCTGLGMQYIKANSKFIGITELE-FACCKKIFPSLAFLESFSGDASNTAPLOPQLO 398
Db 180 KACDGIIGSL-VSAQTVSDSSNIDKFNCTKINGNLFLVTGHHGDPYHTAATNPEKN 238
QY 399 VFETLEITGYLISAMPDSLPLDVSFQNLQVIRGLIHNGAVSLTQGLISWLGRL 458
Db 239 IFQYREITGYLISQSWEMNTDFRVSNTLVITGRLVSLSLIKQGITSLQFQSL 298
QY 459 RELSGSLALHNNHLCFRTVPMDQFRNPHQLHTANPEDECVGEGLAQQLCARG 518
Db 299 KQISAGNIYITDNSMLCYHTVNTSLFTSPQKTVIHRKKAENCTADGVCNELSSD 358
QY 519 HGWGPTQCVNCSOFLRGCEVCECRYLQGLPREYVNAHCLCPHECOP-QNGSVTCF 577
Db 359 GCMGPRDQCLSKRFIRGTCISCNLYDEFRFPANGSCMCDQCEMENDMTTCY 418
QY 578 GPEADQCAAHYKPPFCVACPSGVKPDLSYPMKWPDEBACOPCPINTCTHSCVDL 637
Db 419 GPCPDHCTKCHFPKCGPCVCKCDGLOGANSF--IFKYADEDECHPCHPNCTQGRGP 476
QY 638 DDGC-----PAEGRASPLTISVAVV-GILLVVLGVVGLILRRQCKRKY 685
Db 477 ASHDIYPMTRQSLPQPHAR-TPL--IAAGVIGGLFIVMGLTFAVYARKSK-KKR 532
QY 686 TMRRLQETELVEPLTPSGAMPNOAKMLKEFNFTVSFMLRVPKVASHLETVYKGIW 745
Db 533 ALRRL-ETELVEPLTPSGTAPNOQRLIKE-----TELKRVKVLGSAFGTVYKGIW 585
QY 746 IPDGENVKIPVALIKLRENTSPKANKELIDBAVYVAGSGSYVSRLLGICLTSVQVLTQ 805
Db 586 VPEGETVAKIPVALIKLRENTSPKANKELIDBAVYVAGSGSYVSRLLGICLTSVQVLTQ 805
QY 806 LMPYGLLDHYRENGRLGSODLNMCMQIAKMSYLEDVRLVARDLAARVLYKSPNHV 865
Db 646 LMPHGLLDHYRENGRLGSODLNMCMQIAKMSYLEDVRLVARDLAARVLYKSPNHV 865
QY 866 KITPGLARLIDETEHADGCKVPIKMMALLESILRRPFTHQSDWSYGVTVWELMTFG 925
Db 706 KITPGLARLIDETEHADGCKVPIKMMALLESILRRPFTHQSDWSYGVTVWELMTFG 925
QY 926 AKPYGIPAREIPDLLEKGERLPORPCTIDVYIMVCMIMDSECRPRELSEFSRM 985
Db 766 AKPYGIPAREIPDLLEKGERLPORPCTIDVYIMVCMIMDSECRPRELSEFSRM 985
QY 986 ARDPQRYVIONED-LGPASPLDSTFYSLLEDMDGDLVAEEVLYPQGFPCDDPAPG 1044
Db 866 ARDPQRYVIONED-LGPASPLDSTFYSLLEDMDGDLVAEEVLYPQGFPCDDPAPG 1044
QY 1045 AGWVHHNRSSSTSGGGLTLGLEPSESEAPRS--PLAP-SEGAGSDVDGDLGMAA 1101
Db 885 SRTIDSNRNOFTVZDGGYAAEQV-PMPYRAPGCTIPEAVYAGATATEIFEDCCNMTL 943
QY 1102 KGLQSLPTHDSPLOKRYSEDPVLPSS-----ETDGYVALTCSPPQRYVNOBPVAPQ 1154
Db 944 KQVATLAKEDSSTQRYGADPTVFIPEVIRGELDEDEGYTDMKPKFTDVLNPEVNF 1003
QY 1155 PPSRREGPLPAP-RAGATLBERAKTSLFGKNCVVDVF-----AGCAVENPEYILT 1205

```

```

Db 1004 VSRKNGDLQAVDNEPEYHN-----APNGQKAEDEVNEPLYNTEFANTLENAEL-- 1054
QY 1206 QGGAAPQPPPPAPSPAPDNLVYQDPPERPA--PSTFKCTP-----A 1249
Db 1055 -----KNLPERKAKAFDNPDMNHSLEPSTLQHPYLDQYSTKYFYKNGRIRPIVA 1108
QY 1250 ENPEYL 1255
Db 1109 ENPEYL 1114

```

RESULT 6

```

ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TEMBLrel. 03, last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, last annotation update)
DE Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=91033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9917347; PubMed=1007531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RL rubripes.";
RU Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P1362; 1FCG.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR Prodom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A33039258B647B9 CRC64;

```

Query Match

Best Local Similarity 39.4%; Pred. No. 6.5e-160; Matches 510; Conservative 158; Mismatches 427; Indels 198; Gaps 33;

```

QY 9 WGLLLALPP--GAASDQ-----VCTGTDKMLRLPASPEHLMDLRLHYGQCYVQGNLEL 62
Db 4 WRLIMCVASRLBAASQDQEAQVCPQTQNGLSGQENQYLNKXRYGCEIIMGNLEI 63
QY 63 TYPTNASLSPLODIOEVQYVLIANNQRYVQLRLRYGQLFEDNYALAVLDMGDP 122
Db 64 TQIESWDSFLLTITIEVGYLIANNHFOELPLGGLRIYRGLSLERFALSVFLN-- 120
QY 123 LNNTPVATGAPGGLRELOLRSLTEILKGVLIQRPOLCYQDTILMKDIFHKNQALAT 182
Db 121 ----YKDG--PSGLNQLGIMNTLTLIDGVOQINKKYLRYGFWYVWRDI--RNNAPLE 173
QY 183 LIDTNRSPACHPSPCKSGRCWGSSEDCQSLRTTVACGC-ARCKGSLPFDCCHEGCA 241
Db 174 IOFNGRGVCH--KSC-GNYCWPQKDCQILITKTVCAPOCNDRCFGSPRCCHIECA 229
QY 242 AGCTGPHSGCLACLPFNHNSGICELHCPALVTYNTDTFSPMPREGRYTFGASCVTAPCY 301
Db 230 AGCKGSLDIDCFACRLPNDGACVPCQPTLLYKQTFQWETNPNAKYQGSICVQCFT 289

```

```

QY 302 NYLSTVSGCTLVCLPLHNOEYTAEDGTOR-CEKSKPCARVCYGLGMOYIKANSKFIGIT 360
DB 290 HFV-VDGSSCVSVCPKMEY--ERGSQRCCLSGCPCPKCEGTGAE-----QROYDSS 342
QY 361 ELE-FAGCKIFGSLALPESPDPPASNTAPLOPELOVETLEETGLYISAMPDSL 419
DB 343 NIDFINCTKIQSLHPLVLTGILDDKRVNPPDLAKLEFVFRVETLDTLNTQSWREL 402
QY 420 PDLSEFONLQVIRGRILHNGAYSLLTQGLISWLGSLNLSLELSSGLALIHNTLHLCFVHT 479
DB 403 NDLSEFSLTITQGRSLFKRPSLWVMIPTLISGLRSLLEISGSYIISQMLCHYHT 462
QY 480 VPMQDLFRNPH-QALLTANRPEDECVGEGILACQOLARHCWCPGTQCVNCSQFLRQ 538
DB 463 VNMQLFRGRVRANSUNSNRPMACVADRVCDPLSDSCWCPGPDQCLSRNYSRHG 522
QY 539 ECEVECRVLOGLPREYVNAH-CLPCHPECOPOGNSVTCGFEADOCVACAHYKDPFCV 597
DB 523 TVVAGCHFNNSIPREFAGLNGVCVACHPECKPOTGKASCTGPGADECMACTKTRDQPYCM 582
QY 598 ARCPGVKPDLSYMPWKPFDEGACQPCPINTCHSCVDLDDKCPAEGASPLTISVA 657
DB 583 SSCPAGVN-DEKGLIFKFPREHCEPCQONCTOGSCSGPLNDC--LEAARLTSSGQ 638
QY 658 VVGLLVVVGVP-----GLIKRQCKIRKRYTRRLLOETELVEPLTPGAMNQA 710
DB 639 ITGIALGVPAGLICLVLPFLGMLYHNGLAIRRRGARRVLESSESEPLGP-GERKTYV 697
QY 711 QMRILKEFNFTVSFVLPVPSASHLETYYKGIWIDGNVMIKPAIKLRENTSPKAN 770
DB 698 HARILKP-----SDLKIKPILSGVGTGSKGWIPEGVTKIPVAKITIQSSSGRTF 751
QY 771 KEILDEAYVAVGVSPIVSRILGICLTSTVQVLTQMLPYGCLLDHRENGRIGSQDLN 830
DB 752 TEITDHLISGSDHPYVIRLIGCPOTCQLVLTQSSHSLSEHNRQHTSIDPRLN 811
QY 831 MCQIAGKMSYLEVDRVLRHDLAARNLVKSPNHVKTITDGLARLIDITEFHADGKV 890
DB 812 MCQIAGKMSYLEHHRVYHKNLAARNILKNDYQVQSDGVADLYPDDKTYVSETKT 871
QY 891 PIKMALESILRRRFTHQSDVWSYGVTVWELMTFGAXPYDGIPIAREIPDLLEKGERLPO 950
DB 872 PIKMALESILFRYTHQSDVWSYGVTVWEMSGFAPYASVOQEVPSVLEKGERLSQ 911
QY 951 PICTIDYVMIMVCKMWDSCORPRELVEFSRMDRPORFVIOQEDIGPASPLDSTF 1010
DB 932 AICTIDYVMIMVCKMWDENIRPTFKELASDFTRMADPPRYVILIRMEG----- 980
QY 1011 YRSLLEDDMGDLVDAEYLVPOGFCPPDAPAGGVVHRRSSSTRSGGDLTLGLE 1070
DB 981 -----EDSGMGEFL-----RRGSR-----GLLEADLE 1003
QY 1071 PSEBEARSPLABEGAGSDVFDGDLGMC---AAKGLQSLPTHDPRLQ-----R 1117
DB 1004 EDEEE-----GLGDRFATPSLQPSPSWSTSPQSYNSYMWTOUR 1042
QY 1118 YSEDPVPLPSETDGVVAPLTCSPQ- EYVNC-----PDVAPQPPREGE 1162
DB 1043 YD-----FAVSGGHIQVLPMSBPYDITRQLMWQSRSLSVRTLPDRSAFRSSREAE 1096
QY 1163 L--PAAPAGATLERAKTLPKGNKVGVYVAFQGAVENTEYVLTPOGGAAPORHPAPAS 1220
DB 1097 LCEBDGACAGIFRVR-----FGSERGN-----POG----- 1122
QY 1221 PAFDNLVYWDQDPERGAPSTFGKFTAEPS 1253
DB 1123 -----QORKLSTASSPSSFKTMADEDE 1146

```

RESULT 7
Q9BIH9
ID Q9BIH9
AC Q9BIH9; PRELIMINARY; PRT; 1433 AA.

```

DT 01-JUN-2001 (Tremblrel, 17, Created)
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE Putative epidermal growth factor receptor (fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]_TaxID=7165;
RP SEQUENCE FROM N.A.
RC STRAIN=SLA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBD databases.
DR EMBL: AJ301655; CAC35008.1; -.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000345; CysC_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR01245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; Recep_L_domain; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fv; 7.
DR SMART: SM00220; S_TKc; 1.
DR SMART: SM00219; TyKc; 1.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KM Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 28.6%; Score 1959.5; DB 5; Length 1433;
Best Local Similarity 31.9%; Pred No 1.3e-139;
Matches 464; Conservative 197; Mismatches 401; Indels 391; Gaps 38;

QY 26 CCGTDMKRLRPAPEPTHLDMLRHLKYQCCQVQGNLELYLPTNASSLFLQDIQEVQYVL 85
DB 1 CIGTGRMSVFPANREHYHNLDRYNTCTYVDGNLEITWQNTDLNLFQHIREVTGYVL 60
QY 86 IAHNQVROVPLORLIVGTOF-----EDNYALAVLNDGDLNNTPTVTSFGGLREL 140
DB 61 ISLYDPQVILPRLOIRKRTTFKLKMEAYGLFV-----SFSHMTL 104
QY 141 QURSLTEILKGVALLIQRNPOLCYODTILMKDI-FHKNNOLATLLIDNRSRACHPSPMC 199
DB 105 ELPLRDLIGSGVGFNNYNLCHMKSIMBEILLAPQSMQVTFNPSBERCPCHSC 164
QY 200 KSRCKWGESSEDCQSLTRTVACGGA--RCKAPLPTDCHGCAAGCTGPKNSDGLACH 257
DB 165 EVG-CWGEAHNQCQPSKNSKNSPOCSQRCFCGPKRECHJFCAGGCTGPTQSDGLACKN 223
QY 256 FHNSGICEHLCPALTYNTDTPESMNPNEGRTFGASCVTACPYVYSTDVSGCTLVPEL 317
DB 224 FYDDGVCQCECPMGIYNTPTNFWEPNDDGKAYATCYRCKP-EHLKNGACVYRKPK 282
QY 318 HNGEYTAEDGTORCEKSPCARVCYGLGMOYIKANSKFIGITELFAGCKKIFGSLAFL 377
DB 283 GMPQNSE-----CVPCKGVCPKTCPGEGI-----VHSDNIG-----NYKDCIIEGSEIL 329
QY 378 PESPDGPASNT-----APLOEOLQVETLEELIGYLYISAMPSLDLISVFOMLQ 429
DB 330 DQSFQGVQVYTNFSFGPRYIKIDPRLEVFSTVAKEITGFNIOAHHPNFTTLNFRNLE 389

```


RN [1] *
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC027080; AA027080.1;
 KW Hypothetical protein.
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E3101B0 CRC64;
 Query Match 25.4%; Score 1739; DB 11; Length 367;
 Best Local Similarity 88.0%; Pred. No. 9.5e-124;
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;
 QY 895 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICT 954
 DB 1 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICT 60
 QY 955 IDVYIMVYKCMWIDSECRPRFRELVESEFRMAADPQRFVYIQNEDLGASPLDSTFYRSL 1014
 DB 61 IDVYIMVYKCMWIDSECRPRFRELVESEFRMAADPQRFVYIQNEDLGASPLDSTFYRSL 120
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 1074
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 180
 QY 1075 EARRSLPABEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 1134
 DB 181 EARRSLPABEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 240
 QY 1135 APLTCSPOPEYVNPQDVRRPQPSRPGPLPAARPAATLERAATLSPGKNGVYKDVAFG 1194
 DB 241 APLTCSPOPEYVNPQDVRRPQPSRPGPLPAARPAATLERAATLSPGKNGVYKDVAFG 300
 QY 1195 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQDPFRGAPPSFTFKTPTAENPEY 1254
 DB 301 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQDPFRGAPPSFTFKTPTAENPEY 360
 QY 1255 LGLDVFPV 1261
 DB 361 LGLDVFPV 367
 RESULT 10
 Q8WYV0 PRELIMINARY; PRT; 412 AA.
 AC Q8WYV0;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical 44.7 kDa protein.
 GN PP3659.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
 RA Wan D.F., Gu J.R.;
 RT "Novel human cDNA clones with function of inhibiting cancer cell
 RT growth."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF318349; AAL55856.1;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00018; EF_HAND; UNKNOWN 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397EFF3F27D2BC CRC64;
 Query Match 24.8%; Score 1697.5; DB 4; Length 412;
 Best Local Similarity 80.5%; Pred. No. 1.6e-120;
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;
 QY 895 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICT 954
 DB 1 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPPICT 60
 QY 955 IDVYIMVYKCMWIDSECRPRFRELVESEFRMAADPQRFVYIQNEDLGASPLDSTFYRSL 1014
 DB 61 IDVYIMVYKCMWIDSECRPRFRELVESEFRMAADPQRFVYIQNEDLGASPLDSTFYRSL 120
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 1074
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 180
 QY 1075 EARRSLPABEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 1134
 DB 181 EARRSLPABEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 240
 QY 1135 APLTCSPOPEYVNPQDVRRPQPSRPGPLPAARPAATLERAATLSPGKNGVYKDVAFG 1194
 DB 241 APLTCSPOPEYVNPQDVRRPQPSRPGPLPAARPAATLERAATLSPGKNGVYKDVAFG 300
 QY 1195 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQDPFRGAPPSFTFKTPTAENPEY 1254
 DB 301 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQDPFRGAPPSFTFKTPTAENPEY 360
 QY 1227 YWMD-QDPPER-----GAPSTFGCTPAEN 1251
 DB 361 YWMD-QDPPER-----GAPSTFGCTPAEN 410
 RESULT 11
 Q86712 PRELIMINARY; PRT; 729 AA.
 AC Q86712;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Polypeptide.
 GN POLYPEPTIDE.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 CX NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities."
 RL Oncogene 9:1307-1320(1994).
 DR EMBL: S69372; AAC60725.1;
 DR HSP: P03322; 1A6S.
 DR InterPro: IPR007719; Euk_pkinase.
 DR InterPro: IPR004028; Retro_M.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF02813; Retro_M; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 729 AA; 80649 MW; 84DF6914E91D63 CRC64;
 Query Match 23.9%; Score 1639; DB 15; Length 729;
 Best Local Similarity 52.5%; Pred. No. 1e-115;

Matches 346; Conservative 79; Mismatches 126; Indels 108; Gaps 16;

QY 569 PQNSVTCGPEADCCVCAAHKDPFCVACSPGVKRDLSYMPIMKPEDEGACQCPPI 628
 DB 141 PEETATPKTGP--DHCKMCAHFIDGPHCVKACPAVLGENDTL-VMKADANAVALCHP 197
 QY 629 NCTHSCVDLDDKGCFAEORASPLTSIVAVV-GILLVVLGVFGLIKRQOKIRKXTM 687
 DB 198 NCTRCCKGFGLEGCE---NGSKTPIAAGVVGGLLVVGLIGLYLRR-HIVRKXTL 253

QY 668 RLLDETEIVEPLTSSGAMPNOQMRILKEFNFTVSWLVPKYSASHLETYVKGIMIP 747
 DB 254 RLLDETEIVEPLTSSGAMPNOQMRILKEFNFTVSWLVPKYSASHLETYVKGIMIP 307

QY 748 DGENYKIPVAIKVLRNTPSPKANKELDEAVYMAVGSPVYSRLGLISTVQLVTQIM 807
 DB 308 ESEKVKIPVAIKVLRNTPSPKANKELDEAVYMAVGSPVYSRLGLISTVQLVTQIM 367

QY 808 PYGCLLDVHVRNKGRLSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKI 867
 DB 368 PYGCLLDVHVRNKGRLSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKI 427

QY 868 TDFGLARLLDIDETRYHADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTGAX 927
 DB 428 TDFGLARLLDIDETRYHADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTGAX 487

QY 928 PYDGIPIAEIPDLLEKGERLPQPICTIDVYIMVYKMMIDSECRPREFELVSEFSMAR 987
 DB 488 PYDGIPIAEIPDLLEKGERLPQPICTIDVYIMVYKMMIDSECRPREFELVSEFSMAR 547

QY 988 DQRFVIVQ-NEEDGAPASPLDSTFYRSLLEDMDMDVDAEYLVPOGFFCPRPAPGAG 1046
 DB 548 DQRFVIVQ-NEEDGAPASPLDSTFYRSLLEDMDMDVDAEYLVPOGFFCPRPAPGAG 598

QY 1047 GMVHHRHRSSTRSOGGDLTLGLPESEEAAPRSP-APSEGAGSDVPDGLMGMAA 1101
 DB 599 -----NSPST-----SRPLLSLSATSNNSATNCID----- 625

QY 1102 KGLQSLPHDPSPLORVSEDPVPLPSET--DGVAAPLTGSPQPEYVNOQPVPRQPSR 1159
 DB 626 KNGQGHVPRDSEFVYKYSDDPGNLFLESIDGFL-----PABEYVNO-Q-LMPKPS-- 675

QY 1160 EGPLPAPAPAGATLERAKTLSPKNGVVKVDF-----AFGAVENPEYL 1203
 DB 676 -----TAMVQNIINNISLTAISKLPMSDRYQNSHSTAVIDNPEYL 715

RESULT 12

Q86714 PRELIMINARY; PRT; 567 AA.

AC Q86714;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE V-erbB protein (Fragment).
 GN V-erbB.
 OS Avian rous-associated virus type 1.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 NCBI_TaxID=11950;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94203659; PubMed=8152791;
 RA Venström B., Raynosccheck C., Jansson L., Doederlein G., Lhotak V.,
 RA Johnson A., Beug H.;
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with
 RT different transforming capacities.";
 RL Oncogene 9:1307-1320(1994).
 DR EMBL; S69372; AAC60727.1; .
 DR HSSP; F11362; IFGK.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00069; kinase; 1.

DR PRINTS; PR00109; TYRKINASE.
 DR Prodom; PD000001; Euk_kinase; 1.
 DR SMART; SM00219; Tyrc1.1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KM Tyrosine-protein kinase.
 FT NON TER
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 23.9%; Score 1637; DB 15; Length 567;
 Best Local Similarity 53.1%; Pred. No. 1e-115;
 Matches 345; Conservative 77; Mismatches 120; Indels 108; Gaps 16;

QY 578 GPBADCVCACAHKDPFCVACSPGVKRDLSYMPIMKPEDEGACQCPINCTHSCVDL 637
 DB 1 GP--DHCKMCAHFIDGPHCVKACPAVLGENDTL-VMKADANAVALCHPNTRCCKG 57

QY 638 DDKGCEAPORASPLTSIVAVV-GILLVVLGVFGLIKRQOKIRKXTMRLLOETEL 696
 DB 58 GLEGCEP---NGSKTPIAAGVVGGLLVVGLIGLYLRR-HIVRKXTLRLLOREL 113

QY 697 VEBLTSGAMPNOQMRILKEFNFTVSWLVPKYSASHLETYVKGIMIPQGENYKIPV 756
 DB 114 VEBLTSGAMPNOQMRILKEFNFTVSWLVPKYSASHLETYVKGIMIPQGENYKIPV 167

QY 757 AIKVLRENTSPKANKELDEAVYMAVGSPVYSRLGLISTVQLVTQIMPYGCLLDHV 816
 DB 168 AIKVLRENTSPKANKELDEAVYMAVGSPVYSRLGLISTVQLVTQIMPYGCLLDHV 227

QY 817 RERKRGSLSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLAKL 876
 DB 228 RERKRGSLSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLAKL 287

QY 877 DIDETEYHADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAE 936
 DB 288 GADKEXYHADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAE 347

QY 937 IPDLLEKGERLPQPICTIDVYIMVYKMMIDSECRPREFELVSEFSMARDPQRFVIVQ 996
 DB 348 ISSVLEKGERLPQPICTIDVYIMVYKMMIDSECRPREFELVSEFSMARDPQRFVIVQ 407

QY 997 -NEEDGAPASPLDSTFYRSLLEDMDMDVDAEYLVPOGFFCPRPAPGAGGMVHHRHS 1055
 DB 408 GDERMHLPSFTDSKFTRTLMEEEDMDIVDAEYLVPOGFF-----NS 451

QY 1056 SSTRSGGDLTLGLPESEEAAPRSP-APSEGAGSDVPDGLMGMAKGLQSLPTH 1110
 DB 452 PST-----SRPLLSLSATSNNSATNCID-----RNGQGHVPR 485

QY 1111 DPEPLORYSEDPVPLPSET--DGVAAPLTGSPQPEYVNOQPVPRQPSRPFPLPAPR 1168
 DB 486 EDSEFVYKYSDDPGNLFLESIDGFL-----PABEYVNO-Q-LMPKPS----- 526

QY 1169 AGATLERAKTLSPKNGVVKVDF-----AFGAVENPEYL 1203
 DB 527 -----TAMVQNIINNISLTAISKLPMSDRYQNSHSTAVIDNPEYL 566

RESULT 13

Q64895 PRELIMINARY; PRT; 962 AA.

AC Q64895;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Gag-v-erb-A,v-erb-B protein.
 GN Gag-v-erb-A,v-erb-B.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
 NCBI_TaxID=11661;
 RN [1]
 RP SEQUENCE FROM N.A.

RA MEDLINE=90206603; PubMed=1969616;
 RA Brustin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
 RT "Six amino acids from the retroviral gene gag greatly enhance the
 RT transforming potential of the oncogene v-erb-B";
 RL Oncogene 5:15-24 (1990).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
 DR EMBL: X52209; CA36459.1; -.
 DR EMBL: X52211; CA36459.1; JOINED.
 DR HSSP: P10828; NMLL.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000535; Hormone_rec_119.
 DR InterPro: IPR001723; Sterhmn_receptor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR InterPro: IPR001628; Znf_C4steroid.
 DR Pfam: PF00104; hormone_rec; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00105; zfc4; 1.
 DR PRINTS: PR00398; STRDHOMONER.
 DR PRINTS: PR00047; STROIDFINGER.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR ProDom: PD000035; Znf_C4steroid; 1.
 DR SMART: SM00430; HOL1; 1.
 DR SMART: SM00219; TYK1; 1.
 DR SMART: SM00399; Znf_C4; 1.
 DR PROSITE: PS00031; NUCLEAR RECEPTOR; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; DNA-binding; Nuclear protein; Receptor;
 KW Transcription regulation; Tyrosine-protein kinase;
 KW Zinc-finger.
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B55CE CR664;
 Query Match 23.0%; Score 1575.5; DB 15; Length 962;
 Best Local Similarity 49.7%; Pred. No. 1,1e-110;
 Matches 347; Conservative 76; Mismatches 150; Indels 125; Gaps 19;
 QY 541 VEECRVLOGLPRE-YVNAR-HCLP-----CHPEQ 568
 DB 354 IEKQESYLLAFEHYINRKHNI-PHWSKILMKVADLRMGAVHNRFLMKYECETLS 413
 QY 569 PONGSVTCGPEADQVCAHYKDPFCVACRCPSPGKPLSYMPIKFPDEGACQPCPI 628
 DB 414 PGE-----VGP--DHCKKCAHFIDGPHCVACPAVIGENDTL-VMKYADANAVCQLGHP 465
 QY 629 NCTHSCVDLDKCGPABQASPLTISVAVV-GILLVVLGVVFGILLRQOKIRKXTM 687
 DB 466 NCTRGCKGPELEGP---NGSKTPSIAGVGGLLCLLVVVGIGLYLRRR-HIVKRTLL 521
 QY 688 RLLOETELVEPLTPSGAMNQAMRILKEFNFTVSFMLRVKVSASHLETYKGIWIP 747
 DB 522 RLLORELEVEPLTPSGEAPNQAHRLIKE-----TEFKKVVVLGFGAFGLVYKGLWIP 575
 QY 748 DGNVNVIPVAKILRENTSPKANKELIDEAYYVAGVSPYVRLGLCLTSTVOLVQLM 807
 DB 576 EGKQVTPVAKILRENTSPKANKELIDEAYYVAGVSPYVRLGLCLTSTVOLVQLM 807
 QY 808 PYGLDLHVENGRGSDQLNMCQIAKGSYLEDVRLVHDLAARVNLVKSPPHVKI 867
 DB 636 PYGLDLHVENGRGSDQLNMCQIAKGSYLEDVRLVHDLAARVNLVKSPPHVKI 867
 QY 868 TDFGLARLLDIDETEHADGKVPFKMALESILRRRFTGSDVMSYGVTVWELMTFGAK 927
 DB 696 TDFGLARLLDIDETEHADGKVPFKMALESILRRRFTGSDVMSYGVTVWELMTFGSK 927
 QY 928 PYGIPAREIPDLLEKGERLPOPICTIVYIMVYCMWIDECRPFELVSEFSMAR 987
 DB 756 PYGIPAREIPDLLEKGERLPOPICTIVYIMVYCMWIDECRPFELVSEFSMAR 987
 QY 988 DPGFVYVIO-NEDLGASPLDSTFYVSLLEDMDMDLVDAEYLVPOQGFPCDPDPAFG 1046
 DB 816 DPGFVYVIO-NEDLGASPLDSTFYVSLLEDMDMDLVDAEYLVPOQGFPCDPDPAFG 1046

QY 1047 GWHHRHSSSTRSGGCDLLGLPESEEARSPPLAPSEGASDVFDGDLGGAAGLQS 1106
 DB 867 -----NSPST-----SRTPLSLLSATSN-----NSATKICDR 894
 QY 1107 LPTHPDPLQRYSEDPVPLBSETGYVAPLTCSPQEPYVNOVDVROPSPREGLPAA 1166
 DB 895 NGGH-----PYREDGFL-----PAPRYVNO-LMPKPSYAWVQNOIY 930
 QY 1167 RPAGAT-LEBAKTLSPKNGVYKDVAFAGAVENDEYL 1203
 DB 931 NYSILTAISKLPMSDRYQN-----SSTAVDEEYL 961
 RESULT 14
 Q85468 PRELIMINARY; PRT; 545 AA.
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DE 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Avian erythroblastosis virus (Ts34) v-erbB gene.
 OS Avian erythroblastosis virus.
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_Taxid=11861;
 RX SEQUENCE FROM N.A.
 RX MEDLINE=88217326; PubMed=2897102;
 RA Scotting P., Venstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
 RT "Common site of mutation in the erbB gene of avian erythroblastosis
 RT virus mutants that are temperature sensitive for transformation.";
 RL Oncogene Res. 1:265-278(1987).
 DR EMBL: X06943; CA30024.1; -.
 DR HSSP: P11362; IFGK.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00219; TYK1; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR ATP-binding; Tyrosine-protein kinase.
 KW Tyrosine-protein kinase.
 SQ SEQUENCE 545 AA; 60899 MW; 140DC8C8CA08FAF4 CR664;
 Query Match 22.9%; Score 1567; DB 15; Length 545;
 Best Local Similarity 52.7%; Pred. No. 2e-110;
 Matches 334; Conservative 73; Mismatches 129; Indels 98; Gaps 16;
 QY 578 GPEADQVCAHYKDPFCVACRCPSPGKPLSYMPIKFPDEGACQPCPINCHSCVDL 637
 DB 1 GP--DHCKKCAHFIDGPHCVACPAVIGENDTL-VMKYADANAVCQLCHPCTRGCKGP 57
 QY 638 DDKGCPABQASPLTISVAVV-GILLVVLGVVFGILLRQOKIRKXTMRLLOETEL 696
 DB 58 GLEGP---NGSKTPSIAGVGGLLCLLVVVGIGLYLRRR-HIVKRTLLRLOEEL 113
 QY 697 VEPLTPSGAMNQAMRILKEFNFTVSFMLRVKVSASHLETYKGIWIDGENVKIPV 756
 DB 114 VEPLTPSGEAPNQAHRLIKE-----TEFKKVVVLGFGAFGLVYKGLIPEGKVTIPV 167
 QY 757 AIKVLRENTSPKANKELIDEAYYVAGVSPYVRLGLCLTSTVOLVQLMPYGLLDHV 816
 DB 168 AIKELRENTSPKANKELIDEAYYVAGVSPYVRLGLCLTSTVOLVQLMPYGLLDYI 227
 QY 817 RENNGRLGSDQLNMCQIAKGSYLEDVRLVHDLAARVNLVKSPPHVKITDFGLARLL 876
 DB 228 RENHNDIGSOYLLNMCQIAKGSYLEDVRLVHDLAARVNLVKSPPHVKITDFGLARLL 876
 QY 877 DIDETEHADGKVPFKMALESILRRRFTGSDVMSYGVTVWELMTFGAKPYGIPARE 936
 DB 288 GADKEVTHABGKVPFKMALESILRRRFTGSDVMSYGVTVWELMTFGSKPYDGIPIASE 936

QY 937 IPDLKGERLPQPICTIDVYIMVCMWIDSECRPRELVESEFARMADPQRFVYIQ 996
 DB 348 ISSVLEKGERLPQPICTIDVYIMVCMWIDSECRPRELVESEFARMADPQRFVYIQ 407
 QY 997 -NEULGPASPLDSTFFISLLEDDMDGLVDAEELVVOQGFCCDPAPGAGVYHHRHS 1055
 DB 408 GDERMHLPSPTDSKFYRTLMEEDEMDIVDAEVLVPHQGF-----NS 451
 QY 1056 SSTRSGGDLTLGLEPSEEPASPL-----APEGAGDVFGDGLGMGAAGLQSLPTH 1110
 DB 452 PST-----SRTPLLSLSLATSNMNTATNIDNG-----H 481
 QY 1111 DPSPLORYSEDPVPLBENDGYAPLTCSPQFVYVQPDVPPQSPSPREGPLPAAPAG 1170
 DB 482 -----PYREDGFL-----PAPRYVQ--LMPKPSSTAMVOIQIYVYIS 517
 QY 1171 AT-LEPAKTLSPGKGVVQVDFAFGAGVENEPEYL 1203
 DB 518 LTAISKLPMSRYON-----SHSTAVDNEPEYL 544

RESULT 15

Q9WVF5 PRELIMINARY: PRT: 655 AA.

AC Q9WVF5.
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
 GN EGFR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER.
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Mahle N.J.;
 RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA; TISSUE=LIVER.
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Mahle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=LIVER.
 RX MEDLINE=21085660; Pubmed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa K., Hara A., Fukunishi Y., Komori H., Adachi J., Fukuda S., Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kato K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Flischnmann W., Gaasterland T., Gissi C., King B., Kochwa H., Kusl P., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J., Schirai L.M., Staudt F., Suzuki R., Tomita M., Wagner D., Washio T., Sasaki K., Okido T., Futuno M., Kono H., Baldarelli R., Barsh G., Blake J., Botelli D., Boujunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F., Suzuki H., Togo-Oka K., Wang K.H., Weitz C., Whitteker C., Williams L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayaishiaki Y.;
 RT Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690 (2001).
 DR EMBL: AF124513; AAD44149.1; -
 DR EMBL: AF275366; AAG28047.1; -
 DR EMBL: AF275364; AAG28047.1; JOINED.
 DR EMBL: AF275365; AAG28047.1; JOINED.
 DR EMBL: AK004944; BAB23688.1; -
 DR EMBL: AK004883; BAB23641.1; -
 DR EMBL: AK004911; BAB23662.1; -
 DR MGI: 95294; Egfr.
 DR InterPro: IPR000434; EGFR_L_domain.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF01030; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 3.
 KW Receptor.
 SQ SEQUENCE 655 AA; 72906 MW; 68340631B0C928CB CRC64;

Query Match 22.0%; Score 1506.5; DB 11; Length 655;
 Best Local Similarity 44.5%; Pred. No. 1e-105;
 Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 LIALALPAGAA--STVCTGTDMKRLPASPETHLMDRLHYOGQVWQGNLELYPTN 68
 DB 14 LITLCAAGALBKKVCGGTSNRLTGTFEDHLSLQRMVNNCEVGNLEIITYVGN 73
 QY 69 ASLSFLQDIEVQGYLLAHNOVQVPLORLRIYVGTQLFEDNYALAVLDNDPLNTTP 128
 DB 74 YDLSEFKTIQEVAGVLLALNVERIPLENQIIRGNALYENTVALAILSN----- 124
 QY 129 VTGASPGGLREVLTSLTEILKGVILQNPOLQYDTILMKDI----FKKNQALTLTI 184
 DB 125 -YGTNRTGRLPMLNGLGILGARFSNPIILCMMDITIQMDIYQVNFMSMSDL--- 180
 QY 185 DTRNSBACHPCSPMKGSRFCWGESSEDCSLTRTYCAGCA-RCKGPLPTDCEHQAAG 243
 DB 181 -QSHPSCKPCDPSGPGSCWGGSENCQKLTIKIQAQCSHRGSRSPSDCHNQAAG 239
 QY 244 CGPKRSDCLACLHFNHSGICELHCPALVTYNTDFEEMPNREGYTGASCVTAQPNY 303
 DB 240 CTGPRSDCLVQKQDEATCKOTCPPLMLNPTTYQMDVNEGKISFGATCVKCPRY 299
 QY 304 LSTDVSGCTLVCPILHNOEVTAEQGTQCEKSKPCARVYGLGMQYIRANSKFIGITE 363
 DB 300 VVTDHGSVYRACGPDIYEV-EDGIRKCKCKGCPKVCNGIGIEPK-DTISINATNIK 357
 QY 364 -FAGCKKIGSLAFIPESFDGDPASNTAPLEQQLQVETLEITIGLYLISMPDSDL 422
 DB 358 HFKYCTALSGDHLIPVAFKGSFRTPLDRELEIKTYVEITIGFLIQAMPDNDL 417
 QY 423 SVFONLQVIRGRILHNGAVSLTLOGISLWLSRLSRELGSGLALIHNTHLCHFVPM 482
 DB 418 HAFENDELIRGRTKRGQFSLAVGLNITSLGRLKELISDQDVILISGRNLCYANTINM 477
 QY 483 DOLFNNPHQALITNRPDEBCVGEGLACHQLCARHCWGPPTQVNCSCGLRQCEVE 542
 DB 478 KTLFGFPNOKTIRIMNRAKCKCAVNHVNCPLCSSEGGCWGEFFRDCVNSRRECEV 537
 QY 543 ECRVVGQPREVYVNRHCLPCPCGPOGNGSVTCGPEADQVACAHKDPFCVARGPS 602
 DB 538 KNITIGEREVEVSECIQCHPELPLQANNTTCGRPDNLCQCAHTIDPHCVKTCPA 597
 QY 603 GVKPDLSTYPIWKPDEEGACQPCPINTGSCVLDLDDKGC 642
 DB 598 GIMGENNTL-VKVKADANNVCHLCHANCTYGCAGGGLGQC 636

Search completed: July 22, 2003, 09:24.40
 Job time : 51.0485 secs

Tue Jul 22 12:42:54 2003

seq4-695-709-14.rpt

Page 13

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 22, 2003, 08:41:54 ; Search time 37.1104 Seconds
(without alignments)
4527.811 Million cell updates/sec

Title: SEQ4-695-709-14

Perfect score: 6853
Sequence: 1 MELALCRWGLLALLPPGA.....TFKTPPAENPEYLGLDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: A_Geneseq.101002.*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 24: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	6647	97.0	1255	21 AAY92620
2	6647	97.0	1255	22 AAE12130
3	6647	97.0	1255	22 AAB60167
4	6647	97.0	1255	23 AAU74545
5	6641	96.9	1255	17 AAU01111
6	6641	96.9	1255	21 AAW92406
7	6641	96.9	1255	21 AAB21198
8	6641	96.9	1255	21 AAY84780
9	6641	96.9	1255	22 AAB85458
10	6641	96.9	1255	22 AAG88267

11	6641	96.9	1255	23 AAE24067	Human Her-2 protei
12	6641	96.9	1255	23 AAE20479	Human Her-2/neu pr
13	6641	96.9	1255	23 AAE51143	Human Her-2/neu on
14	6641	96.9	1255	23 AAU77114	Human Her-2/neu po
15	6598	96.3	1433	14 AAR39568	Sequence of c-erbB
16	6477	94.5	1223	14 AAU98923	Human breast cance
17	6324	92.3	1200	21 AAB21199	Human Her-2/neu pr
18	5854.5	85.4	1256	23 AAB21199	Rat Her-2/neu prot
19	5854.5	85.4	1256	23 AAE51144	Rat Her-2/neu onco
20	5829.5	85.1	1256	22 AAB21206	Mouse Her-2/neu pr
21	5829.5	85.1	1256	22 AAG52860	Amino acid sequenc
22	5829.5	85.1	1256	22 AAE51151	Mouse Her-2/neu on
23	4811	70.2	919	23 AAB21203	Human Her-2/neu fu
24	4811	70.2	919	23 AAE51158	Her-2/neu extracel
25	4061.5	59.3	920	23 AAE51152	Mouse Her-2/neu ex
26	4061.5	59.3	926	23 AAE51152	Mouse Her-2/neu ex
27	3695	53.9	712	21 AAB21204	Human Her-2/neu fu
28	3595	53.9	712	23 AAE51149	Her-2/neu extracel
29	3552	51.8	782	18 AAM19764	Her-2/GM-CSF immuno
30	3550	51.8	653	21 AAB21200	Human Her-2/neu on
31	3550	51.8	653	21 AAE51145	Human ErbB2 oncopr
32	3512	51.2	645	22 AAB60408	Human ErbB2 extrac
33	3512	51.2	645	22 AAB61593	DC8ecFv-erbB2EC fu
34	3447	50.3	951	21 AAY44993	Extracellular port
35	3344	48.8	624	11 AAR08222	Human EGF receptor
36	3065	44.7	1210	21 AAB19239	Amino acid sequenc
37	3065	44.7	1210	21 AAY50616	Human EGF receptor
38	3065	44.7	1210	23 AAE23019	Human Her-1 protei
39	3065	44.7	1210	23 AAE50768	Human epidermal gr
40	3063	44.7	1210	22 AAB68400	Amino acid sequenc
41	3053.5	44.6	654	21 AAB21205	Rat Her-2/neu prot
42	3053.5	44.6	654	21 AAE51150	Rat Her-2/neu onco
43	3024	43.1	1210	23 AAB51763	Human epidermal gr
44	2997	43.7	583	23 AAE20483	Human protein for
45	2997	43.7	587	23 AAE20481	Human protein for

ALIGNMENTS

RESULT 1	AAV92620 standard; Protein, 1255 AA.
ID	AAV92620
AC	AAV92620;
XX	10-AUG-2000 (first entry)
DT	Human heregulin 2 (Her2).
XX	Her2; vaccination; cytotoxic T-lymphocyte immunity;
XX	self-protein; cancer; breast cancer; prostate cancer;
KW	cell-associated peptide antigen; foreign epitope.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	Domain
FT	Location/Qualifiers
FT	1..173
FT	/label= "N-terminal"
FT	/note= "mature polypeptide"
FT	5..25
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	59..73
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	103..117
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	149..163
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	174..323
FT	Domain

FT /label= Cysteine_rich_domain
 FT Region
 FT 210..224
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 250..264
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 324..483
 FT /label= Ligand_binding_domain
 FT 325..339
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 369..383
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 465..479
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 484..623
 FT /label= Cysteine_rich_domain
 FT 579..593
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 624..654
 FT /label= Transmembrane_domain
 FT 632..652
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 653..667
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 655..1010
 FT /label= Tyrosine_kinase_domain
 FT 661..675
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 695..709
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Region
 FT 710..730
 FT /label= insertion region
 FT /note= "suitable for foreign epitope insertion"
 FT Domain
 FT 1011..1235
 FT /label= C-terminal_domain
 FT
 PN W0200020027-A2.
 XX
 XX 13-APR-2000.
 XX
 PF 05-OCT-1999; 99MO-DK00525.
 XX
 PR 05-OCT-1998; 98DK-0001261.
 XX 20-OCT-1998; 98US-0105011.
 PA (MEB1-) M & E BIOTECH AS.
 PI Steina L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,
 XX Gautam A, Birk P, Karlsson G;
 XX WPI: 2000-34997/30.
 DR N-PSDB; AAA09455.
 XX
 XX Claim 62; Page 193-198; 220pp; English.
 PT Inducing immune responses to weakly immunogenic, tumor associated
 PT peptide antigens for the treatment of breast and prostate cancer
 XX
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
 CC response. Subdominant CTL epitopes, antibody binding regions and
 CC cysteine residues involved in disulfide bonds are preserved in the
 CC immunogenized forms. Regions suitable for the insertion of foreign T
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic
 CC cell-associated peptide antigens (PA) such as those associated with
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
 CC The method comprises effecting simultaneous presentation by antigen
 CC producing cells (APCs) of the animal's immune system of: (1) at least 1
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
 CC B-cell group derived from the cell-associated PA; and (2) at least 1
 CC first T helper cell group which is foreign to the animal. Analogues of
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
 CC part of all known and predicted CTL and B-cell epitopes of the respective
 CC PA and including at least one foreign T helper epitope are also claimed.
 CC The method is used to treat prostate, prostate/breast or breast cancer
 CC when the PA is human PSM, FGF8b and Her2, respectively.
 CC
 SQ Sequence 1255 AA;

Query Match 97.0%; Score 6647; DB 21; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;

QY 1 METALCRWGLLLALPPGAASVCTGTDMKRLRPASPEHLDMLRHLVGGCQVQGNL 60
 Db 1 METALCRWGLLLALPPGAASVCTGTDMKRLRPASPEHLDMLRHLVGGCQVQGNL 60
 QY 61 EITVLPNASLFLDIOEVQGVYLIANOVROYPLQRLRVGTQLFEDNYALAVDNG 120
 Db 61 EITVLPNASLFLDIOEVQGVYLIANOVROYPLQRLRVGTQLFEDNYALAVDNG 120
 QY 121 DLNNTPTVTGASPGGLRELQRLSTELTKGVYIQRPQCYOPTIMKQIFHNQOLA 180
 Db 121 DLNNTPTVTGASPGGLRELQRLSTELTKGVYIQRPQCYOPTIMKQIFHNQOLA 180
 QY 121 DLNNTPTVTGASPGGLRELQRLSTELTKGVYIQRPQCYOPTIMKQIFHNQOLA 180
 Db 121 DLNNTPTVTGASPGGLRELQRLSTELTKGVYIQRPQCYOPTIMKQIFHNQOLA 180
 QY 181 LTLIDTNSRACHPCSPCKSCRCWSESSDQSLTRTVACAGCARCKGPPPTDCHEQC 240
 Db 181 LTLIDTNSRACHPCSPCKSCRCWSESSDQSLTRTVACAGCARCKGPPPTDCHEQC 240
 QY 241 AAGCGPRHSDCLALHNHNSGICGLHPALVTYVTFEEMPNEGYTGASVTCAP 300
 Db 241 AAGCGPRHSDCLALHNHNSGICGLHPALVTYVTFEEMPNEGYTGASVTCAP 300
 QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVCYGLGMOYIRANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVCYGLGMOYIRANSKFIGIT 360
 QY 361 EIEPAGCKKIFGSLAFPESEFDGDPASNTAPLQEPQLOVFTLEBITGYLYISAMPDSL 420
 Db 361 EIEPAGCKKIFGSLAFPESEFDGDPASNTAPLQEPQLOVFTLEBITGYLYISAMPDSL 420
 QY 421 DLVSFQNTQVIRGRILHNGAVSLTQGGISWLGRLSRLRETSGLALIHNTHLCFVHTV 480
 Db 421 DLVSFQNTQVIRGRILHNGAVSLTQGGISWLGRLSRLRETSGLALIHNTHLCFVHTV 480
 QY 481 PMDQLFRPHQALLHTANRPEDEVCYEGSLAQHQLCARGHGWPGFTQCVCNSQLRGQEC 540
 Db 481 PMDQLFRPHQALLHTANRPEDEVCYEGSLAQHQLCARGHGWPGFTQCVCNSQLRGQEC 540
 QY 541 VEECVLGLPREYVNAHCLPCHECQPOKGSVTCFEPADQYACAHYDPPCVARC 600
 Db 541 VEECVLGLPREYVNAHCLPCHECQPOKGSVTCFEPADQYACAHYDPPCVARC 600
 QY 601 PEGVXPDLSTYMPIMKFPDEEGACQPCPINTHSCVDLDDKCPABGRASPITSIVAVG 660
 Db 601 PEGVXPDLSTYMPIMKFPDEEGACQPCPINTHSCVDLDDKCPABGRASPITSIVAVG 660
 QY 661 ILVVVAVGVGIIILKROQKIRKTYMRLLQETLVEPLTPSGAMPQAOQRIKXENN 720
 Db 661 ILVVVAVGVGIIILKROQKIRKTYMRLLQETLVEPLTPSGAMPQAOQRIKXENN 720
 QY 721 FTVSFWLWVPKXSAS-HLETYVKGIMIDGENVKIPVAKYLRNTSPSKAKELIDEAYV 779
 Db 720 -----LKKVYLSGAGGTYVKGIMIDGENVKIPVAKYLRNTSPSKAKELIDEAYV 773

QY 780 MAGVSPYVSLIGLITSTVQLVQLMYGCLLDHVRNRLGSLQDILLNCMOIAGKM 839
 DB 774 MAGVSPYVSLIGLITSTVQLVQLMYGCLLDHVRNRLGSLQDILLNCMOIAGKM 833
 QY 840 SYLEDVRLVHRDLAANVLYKSPNHVKTDFGLARLDDIDETREYHADGGKVPDKMALES 899
 DB 834 SYLEDVRLVHRDLAANVLYKSPNHVKTDFGLARLDDIDETREYHADGGKVPDKMALES 893
 QY 900 IARRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPQPICTIDVYM 959
 DB 894 IARRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPQPICTIDVYM 953
 QY 960 IMVKCMIIDSECRPRRELVSERSRARDPQRFVVIQNDLGPASPLDSTFRSLLEDD 1019
 DB 954 IMVKCMIIDSECRPRRELVSERSRARDPQRFVVIQNDLGPASPLDSTFRSLLEDD 1013
 QY 1020 MGDLVDAEYLYVPOQGFPCPDPAFGAGVWHHRHSSSTRSGGDLTLGLEPSEEARPS 1079
 DB 1014 MGDLVDAEYLYVPOQGFPCPDPAFGAGVWHHRHSSSTRSGGDLTLGLEPSEEARPS 1073
 QY 1080 PLAPSEAGSDVFDGLGMAAKGLSLFTHPSPLOQYSEDPVPLPSETDGYAAPLTC 1139
 DB 1074 PLAPSEAGSDVFDGLGMAAKGLSLFTHPSPLOQYSEDPVPLPSETDGYAAPLTC 1133
 QY 1140 SPOPEYVNOVDYRPPSPRSGPLPARPAGATLEPAKTLSPGKGVWVDVAFGAVEN 1199
 DB 1134 SPOPEYVNOVDYRPPSPRSGPLPARPAGATLEPAKTLSPGKGVWVDVAFGAVEN 1193
 QY 1200 PEYLTPQGGAAPQHPHPAPFPAFDMLYWMDQDPERGAPESTFKGTPFAENPEYLGLDV 1259
 DB 1194 PEYLTPQGGAAPQHPHPAPFPAFDMLYWMDQDPERGAPESTFKGTPFAENPEYLGLDV 1253
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 2
 ID AAE12130 standard; Protein; 1255 AA.
 AAE12130;
 18-DEC-2001 (first entry)
 Human tyrosine kinase-type receptor, HER-2.
 Therapeutic compound; major histocompatibility complex; vaccine;
 antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
 adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
 antigen presenting cell; human; tyrosine kinase-type receptor.
 Homo sapiens.
 Key Location/Qualifiers
 Region 774..782
 /note="Antigenic epitope"
 WO200156677-A2.
 20-SEP-2001.
 16-MAR-2001; 2001WO-US40328.
 16-MAR-2000; 2000US-0527487.
 (GENZ) GENZYME CORP.
 Nicolette CA;
 WPI; 2001-616284/71.
 DR N-PSDB; AADI9731.
 XX

PT Novel synthetic therapeutic compound for inducing immune response and
 for use in adoptive immunotherapy, has enhanced binding to major
 PT histocompatibility molecules and enhanced immunoregulatory properties
 PT
 PS Claim 4; Page 63-67; 69pp; English.

The invention relates to synthetic therapeutic compounds (antigenic
 peptides) with enhanced binding to major histocompatibility complex
 (MHC) molecules and enhanced immunoregulatory properties relative
 CC to their natural counterparts. Compounds of the invention are useful
 CC for inducing an immune response in a subject and for use in adoptive
 CC immunotherapy. They are useful as components of anti-cancer vaccines
 CC and to expand immune effector cells that are specific for cancers
 CC characterised by expression of the breast cancer antigen, HER-2.
 CC Polynucleotides that encode peptides of the invention are useful as
 CC hybridisation probes and as primers for the detection of genes of
 CC transcripts that are expressed in antigen presenting cells (APCs), to
 CC confirm transduction of polynucleotides into host cells. The present
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
 CC of the invention are designed based on the HER-2 antigenic peptide
 (774-782).

Sequence 1255 AA:

Query Match 97.0%; Score 6647; DB 22; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;

QY 1 MELALCRWGLLLALLPPGAASITVCGTGMKRLPASPETHLMDLHYGCGVQGNL 60
 DB 1 MELALCRWGLLLALLPPGAASITVCGTGMKRLPASPETHLMDLHYGCGVQGNL 60
 QY 61 ELTYLPTNASLSPLODIQEVQGYVLIANQVQYPLQRLIRVGTQLFEDNYALAVDNG 120
 DB 61 ELTYLPTNASLSPLODIQEVQGYVLIANQVQYPLQRLIRVGTQLFEDNYALAVDNG 120
 QY 121 DPLNNTTPVYGASGGIRLEQLRSLTILKGVLIQNPOLCYDITLMDIPIHKNOLA 180
 DB 121 DPLNNTTPVYGASGGIRLEQLRSLTILKGVLIQNPOLCYDITLMDIPIHKNOLA 180
 QY 121 DPLNNTTPVYGASGGIRLEQLRSLTILKGVLIQNPOLCYDITLMDIPIHKNOLA 180
 DB 121 DPLNNTTPVYGASGGIRLEQLRSLTILKGVLIQNPOLCYDITLMDIPIHKNOLA 180
 QY 181 LTLIDTRSRACHPCSPKMGKSRGWSSEDCSLSTRVACAGCARCKPLPTCCHEOC 240
 DB 181 LTLIDTRSRACHPCSPKMGKSRGWSSEDCSLSTRVACAGCARCKPLPTCCHEOC 240
 QY 241 AAGCTGPKHSDCLALHPNHSIGCELCPLALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKHSDCLALHPNHSIGCELCPLALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
 QY 301 YNYLSTVGSCTLVCPHNOEVTAEDGTQRCCKSKRCARVCYGLGWEHLREYAVTSAN 360
 DB 301 YNYLSTVGSCTLVCPHNOEVTAEDGTQRCCKSKRCARVCYGLGWEHLREYAVTSAN 360
 QY 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVFEETLEITGYLYISAMPDPLP 420
 DB 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVFEETLEITGYLYISAMPDPLP 420
 QY 421 DLSVFQNLQVIRGRIILNNGAYSLTLQGLGISWGLRLRELGSGLALIHNTHLCEVHTV 480
 DB 421 DLSVFQNLQVIRGRIILNNGAYSLTLQGLGISWGLRLRELGSGLALIHNTHLCEVHTV 480
 QY 481 PMDQLFANPQALLHTANRPEDECVGGLACHOLCARGHGCMGPPQCVNCSOLFRLQEC 540
 DB 481 PMDQLFANPQALLHTANRPEDECVGGLACHOLCARGHGCMGPPQCVNCSOLFRLQEC 540
 QY 541 VEEGRVLYQGLPREYVNAHCLCPHPECOPNGSVTCGPEADOCVACAHYKDPFCVAC 600
 DB 541 VEEGRVLYQGLPREYVNAHCLCPHPECOPNGSVTCGPEADOCVACAHYKDPFCVAC 600
 QY 601 PSQVYKPLSYMPIMKPEDEGACOPCINCTHS CVLDDGCAEAGASPLTISAVNG 660
 DB 601 PSQVYKPLSYMPIMKPEDEGACOPCINCTHS CVLDDGCAEAGASPLTISAVNG 660

```

QY 661 ILLVVLGVVFGILLIKRROOKIRKRYMRRLLOETELVEPLTPSGAMPNOAQRILKEFN 720
DB 661 ILLVVLGVVFGILLIKRROOKIRKRYMRRLLOETELVEPLTPSGAMPNOAQRILKEFE- 719
QY 721 FTVSFWLVRPKVASAS-HLETYKGIWIPDGENVKIPVAIKVLRENTSPPANKELIDEAVV 779
DB 720 -----LRKVKVIGSGAFGTIVYKGIWIPDGENVKIPVAIKVLRENTSPPANKELIDEAVV 773
QY 780 MAGVGSPPYVRLILGICITSTVQLVQLTQMPYGLLDHYREKRGSLGSDLLNMCQIAKQM 839
DB 774 MAGVGSPPYVRLILGICITSTVQLVQLTQMPYGLLDHYREKRGSLGSDLLNMCQIAKQM 833
QY 840 SYLEDVRLVTRDLAARVVLKSPHVKITDFGLARLLIDETRYHADGCKVPIKXMALES 899
DB 834 SYLEDVRLVTRDLAARVVLKSPHVKITDFGLARLLIDETRYHADGCKVPIKXMALES 893
QY 900 ILRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959
DB 894 ILRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953
QY 960 IMKCMWIDSECRPRELTVSEFSRMAPDQRFVVIQNEDLGASPLDSTFYRSLLEDD 1019
DB 954 IMKCMWIDSECRPRELTVSEFSRMAPDQRFVVIQNEDLGASPLDSTFYRSLLEDD 1013
QY 1020 MCDLVDAEYLVPQCGFFCPDPAAGAGVNHHRSSSTRSGGDLTLGLSESEEPARS 1079
DB 1014 MCDLVDAEYLVPQCGFFCPDPAAGAGVNHHRSSSTRSGGDLTLGLSESEEPARS 1073
QY 1080 PLAPSEAGSDVDGDLGMAAKGLQSLPTHDSPLQRYSEDPVLPSTDTGYVAPLTC 1139
DB 1074 PLAPSEAGSDVDGDLGMAAKGLQSLPTHDSPLQRYSEDPVLPSTDTGYVAPLTC 1133
QY 1140 SPOREYVNOPDVPFPPSPREGPLPAAPAGATLEBAKTLSFGKGVVQDVAFGGAIVEN 1199
DB 1134 SPOREYVNOPDVPFPPSPREGPLPAAPAGATLEBAKTLSFGKGVVQDVAFGGAIVEN 1193
QY 1200 PEYVLTPOGGAPOPHPPAPSPAFDNLVYWDQDPPERGAPOSTFGKPTIANPEYVGLDV 1259
DB 1194 PEYVLTPOGGAPOPHPPAPSPAFDNLVYWDQDPPERGAPOSTFGKPTIANPEYVGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein: 1255 AA.
AC AAB60167;
DT 03-APR-2001 (first entry)
XX
DE HRR2 transgene plasmid construct encoded protein.
XX
KM Human; HRR2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KM antibody.
XX
OS Homo sapiens.
OS Synthetic.
PN MO200100244-A2.
PD 04-JAN-2001.
PF 23-JUN-2000; 2000MO-US17229.
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH) GENENTECH INC.
XX
PI Erickson S, Schwall R;

```

```

XX MPI: 2001-061962/07.
DR N-PSDB; AAF24297.
DR
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 97.0%; Score 6647; DB 22; Length 1255;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;
1 MELALCEWGLLALILPAGASTVCTGTDMKRLPASPEHLMRLHYGCGVYQGNL 60
1 MELALCEWGLLALILPAGASTVCTGTDMKRLPASPEHLMRLHYGCGVYQGNL 60
61 ELTYLPTNASSFLDIDIEVGVYLIANOVQVPLQRLRYRGTOLFEDNYALAVLDNG 120
61 ELTYLPTNASSFLDIDIEVGVYLIANOVQVPLQRLRYRGTOLFEDNYALAVLDNG 120
121 DELNNTPTVTGASPGGLRELOLRSTLEILKGVYLIQRPQLCYODTILMKDIFHNQOLA 180
121 DELNNTPTVTGASPGGLRELOLRSTLEILKGVYLIQRPQLCYODTILMKDIFHNQOLA 180
181 LTLIDTNSRACHPCSPCKSGRCWSSSEDCQSLTRVCAAGGACRCKGPLPTDCHEQC 240
181 LTLIDTNSRACHPCSPCKSGRCWSSSEDCQSLTRVCAAGGACRCKGPLPTDCHEQC 240
241 AAGCTGPRHSOCLAFHNSGICELHCPALVTYNTDFEEMPNPEGRTGASCVTACP 300
241 AAGCTGPRHSOCLAFHNSGICELHCPALVTYNTDFEEMPNPEGRTGASCVTACP 300
301 YNYLSTDVGSCTLVCPFLHNOEYTAEDGTQRCCKSKPCARVYGIQMOYIKANSKFIGIT 360
301 YNYLSTDVGSCTLVCPFLHNOEYTAEDGTQRCCKSKPCARVYGIQMOYIKANSKFIGIT 360
361 ELFPAGCKKIFGSLAFLESFDDGPASTAFLQBPOLQVETLEITGYLYISAMPDSL 420
361 IQEFAGCKKIFGSLAFLESFDDGPASTAFLQBPOLQVETLEITGYLYISAMPDSL 420
421 DLVSFQNLQVIRGRILHNGAYSLTLQGGISMLGLRSRELGSLALIHNTHTLCFVYTV 480
421 DLVSFQNLQVIRGRILHNGAYSLTLQGGISMLGLRSRELGSLALIHNTHTLCFVYTV 480
481 PWDQLFRNPQALLHTANRPEDECVGBGLACHQLCARHCGMPGTQCVNSQPLRGQEC 540
481 PWDQLFRNPQALLHTANRPEDECVGBGLACHQLCARHCGMPGTQCVNSQPLRGQEC 540
541 VEEGRVLOGLPREYNARHCLPCHPECOPOGNSVTCFPEADQCAAHYDPPFCVARG 600
541 VEEGRVLOGLPREYNARHCLPCHPECOPOGNSVTCFPEADQCAAHYDPPFCVARG 600
601 PSGVAPDLSYMPIWKFPEDEGACQPCPINTHSCVDLDKQCPAEOBASPLTISVAVG 660
601 PSGVAPDLSYMPIWKFPEDEGACQPCPINTHSCVDLDKQCPAEOBASPLTISVAVG 660
661 ILLVVLGVVFGILLIKRROOKIRKRYMRRLLOETELVEPLTPSGAMPNOAQRILKEFE- 719
661 ILLVVLGVVFGILLIKRROOKIRKRYMRRLLOETELVEPLTPSGAMPNOAQRILKEFE- 719
721 FTVSFWLVRPKVASAS-HLETYKGIWIPDGENVKIPVAIKVLRENTSPPANKELIDEAVV 779

```

Db 720 -----LRKXVLSGSAFGTVYKGIWIPDGENKIPVAIKVLEBNTSPRANKETLBEAYV 773
 Qy 780 MAGVSPYVSRLIGICTSTVQVLTQMPYGCILLDHVRENKRLGSDLLNMCMQAKGM 839
 Db 774 MAGGSPYVSRLIGICTSTVQVLTQMPYGCILLDHVRENKRLGSDLLNMCMQAKGM 833
 Qy 840 SYEDVRLVHRDLAARVLKSPHVKITDFGLARLLDIBETVHADGKVPKXMALES 899
 Db 834 SYEDVRLVHRDLAARVLKSPHVKITDFGLARLLDIBETVHADGKVPKXMALES 893
 Qy 900 ILRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYM 959
 Db 894 ILRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYM 953
 Qy 960 IMYKCMWIDSECRPFRELVSERSMARDQRFVYVIONEDLGASPLDSTFFRSLIEDDD 1019
 Db 954 IMYKCMWIDSECRPFRELVSERSMARDQRFVYVIONEDLGASPLDSTFFRSLIEDDD 1013
 Qy 1020 MGDLDVAEELVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEARPS 1079
 Db 1014 MGDLDVAEELVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEEARPS 1073
 Qy 1080 PLAPSEGASDVFDGDIQMGAAKGLSLPTHDSPLQRYSEDPVPLPSTGDYAPLTC 1139
 Db 1074 PLAPSEGASDVFDGDIQMGAAKGLSLPTHDSPLQRYSEDPVPLPSTGDYAPLTC 1133
 Qy 1140 SPOPEYVNOPDVRPQPSPREGPPAPAPAGATLEBAKTLSPKXNVVDVAFGAVEN 1199
 Db 1134 SPOPEYVNOPDVRPQPSPREGPPAPAPAGATLEBAKTLSPKXNVVDVAFGAVEN 1193
 Qy 1200 PEYLTPGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPTFKGTPTAENPEYGLDV 1259
 Db 1194 PEYLTPGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPTFKGTPTAENPEYGLDV 1253
 Qy 1260 PV 1261
 Db 1254 PV 1255

RESULT 4
 ID AAU74545
 AAU74545 standard; Protein; 1255 AA.
 AC AAU74545;
 XX
 DT 23-APR-2002 (first entry)
 XX
 DE Human HER2 (ErbB2) polypeptide.
 XX
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
 KW glial disorder; astrocytic disorder; hypochalamic disorder;
 KW glandular disorder; macrophagal disorder; epithelial disorder;
 KW stromal disorder; blastocoelec disorder; inflammatory disorder;
 KW angiogenic disorder; immunological disorder.
 XX
 OS Homo sapiens.
 XX
 PN US2002001587-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 16-MAR-2001; 2001US-0811123.
 XX
 PR 16-MAR-2000; 2000US-189844P.
 PR 05-OCT-2000; 2000US-238327P.
 XX
 PA (ERIC/) ERICKSON S.
 PA (SCHW/) SCHWALL R.
 PA (SLIW/) SLIWOMSKI M.

PI Erickson S, Schwall R, Sliwowski M;
 XX WPI: 2002-163686/21.
 DR N-PSDB; ABR14058.
 XX
 PT Treating tumour characterised by overexpression of epidermal growth
 factor receptor, ErbB or cancer in mammal, comprises administering
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal
 XX
 PS Example 3; Fig 7; 93pp; English.
 CC The invention relates to treating a tumour in a mammal, where the tumour
 CC is characterised by the overexpression of an epidermal growth factor
 CC receptor (ErbB) and does not respond or responds poorly, to treatment
 CC with an anti-ErbB antibody, comprising administering to the mammal an
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
 CC prostate and bladder, preferably breast cancer. The breast cancer is a
 CC metastatic breast cancer or an aggressive form of metastatic breast
 CC cancer which overexpresses ErbB2. The method is also useful for treating
 CC neuronal, glial, astrocytic, hypochalamic, glandular, macrophagal,
 CC epithelial, stromal, blastocoelec, inflammatory, angiogenic and
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)
 CC polypeptide of the invention.
 XX
 SQ Sequence 1255 AA;
 XX
 Query Match 97.0%; Score 6647; DB 23; Length 1255;
 Best Local Similarity 97.5%; Pred. No. 0;
 Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;
 Qy 1 MELALCWMGLLALLPPGAASYCTGTDMKLRFPASBETHLDMRLHYQGCGVYQNL 60
 Db 1 MELALCWMGLLALLPPGAASYCTGTDMKLRFPASBETHLDMRLHYQGCGVYQNL 60
 Qy 61 ELTYLPTNASLSFLDDIOEVQGYVIAHNOYVPLQRLRYRGTLFEDNYALAVLNG 120
 Db 61 ELTYLPTNASLSFLDDIOEVQGYVIAHNOYVPLQRLRYRGTLFEDNYALAVLNG 120
 Qy 121 DELNNTTPTVGASPGGLRELOLSLTELKGVLIQRNPOLCYOQTLIMKDIFFHKNOLA 180
 Db 121 DELNNTTPTVGASPGGLRELOLSLTELKGVLIQRNPOLCYOQTLIMKDIFFHKNOLA 180
 Qy 181 LTLIDTNSRACHPSCSPKCKSRCKGESSSEDCOSLTRVYVAGGACRGRLPTDCCHQC 240
 Db 181 LTLIDTNSRACHPSCSPKCKSRCKGESSSEDCOSLTRVYVAGGACRGRLPTDCCHQC 240
 Qy 241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYVTFEEMPNREGRYTGASCVTACP 300
 Db 241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYVTFEEMPNREGRYTGASCVTACP 300
 Qy 301 YNYLSTVGSCTVCPPLHNOEVLADGTORCEKSKPCARVYCYGLMGIYKANSKFIGIT 360
 Db 301 YNYLSTVGSCTVCPPLHNOEVLADGTORCEKSKPCARVYCYGLMGIYKANSKFIGIT 360
 Qy 361 ELLEFAGCKKRTGSLAFIPSEFSDGPASNTALQSEOLQVETLEITGYLISAMPDLP 420
 Db 361 ELLEFAGCKKRTGSLAFIPSEFSDGPASNTALQSEOLQVETLEITGYLISAMPDLP 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQIGISWLGRLSRLREGSGALIHNTHTCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQIGISWLGRLSRLREGSGALIHNTHTCFVHTV 480
 Qy 481 PMDQLFRPHQALHTANRPEDCEVSGELACHQICAGHCHGPPPTCCVNCOSPLRQDEC 540
 Db 481 PMDQLFRPHQALHTANRPEDCEVSGELACHQICAGHCHGPPPTCCVNCOSPLRQDEC 540
 Qy 541 VEECRVILQGLPREYVNAHCLPCHPECPQNGSVTCGFPEDQCVACAHYKDPFCVARC 600
 Db 541 VEECRVILQGLPREYVNAHCLPCHPECPQNGSVTCGFPEDQCVACAHYKDPFCVARC 600
 Qy 601 PSSGVKPLDSTYMPKPFDEBGAQCPPCINCTHSCVDLDDKGCAPRAPSPLTISVAVG 660

Db 601 PSQVAPDLSTYMPWKPFDEEGACQCPINCHSCVDDLDKCCPAEQASPLTISVAVG 660
 Qy 661 ILLVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMNQOMRLKEFNN 720
 Db 661 ILLVVLGVVFGILLIKRQOKIRKYTMRLLOETELVEPLTPSGAMNQOMRLKEFNN 719
 Qy 721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKI PAIKYLRENTS PKANKELIDEAYV 779
 Db 720 -----LRKVYKLSGAFGTVKGIWIPDGENVKI PAIKYLRENTS PKANKELIDEAYV 773
 Qy 780 MAGVSPVYSLGLICTLSTVQVLTOLMPYGCILLDHRNRRGRSGSODLLMCMQIAKGM 839
 Db 774 MAGVSPVYSLGLICTLSTVQVLTOLMPYGCILLDHRNRRGRSGSODLLMCMQIAKGM 833
 Qy 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEHADGKVPYIKMMALES 899
 Db 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDGLARLLDIDETEHADGKVPYIKMMALES 893
 Qy 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PARISIPDLEKGERLPPOPTCTIDVYM 959
 Db 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PARISIPDLEKGERLPPOPTCTIDVYM 953
 Qy 960 IMVKCMWIDSECRFRELVSSEFSRMARDPQRFVYIQNEDLGSPASPLDSTFFYSLLIEDD 1019
 Db 954 IMVKCMWIDSECRFRELVSSEFSRMARDPQRFVYIQNEDLGSPASPLDSTFFYSLLIEDD 1013
 Qy 1020 MGDLDVAEEYLVPOQGFCDPAPAGAGMHHRRSSSTSGGGLTLGLPSEEEAPRS 1079
 Db 1014 MGDLDVAEEYLVPOQGFCDPAPAGAGMHHRRSSSTSGGGLTLGLPSEEEAPRS 1073
 Qy 1080 PLASEGAGSDVFGDGLGMAKAGLOSLPTHDPSPQLRYSEDDPVPLPSESTDVYVAPLTC 1139
 Db 1074 PLASEGAGSDVFGDGLGMAKAGLOSLPTHDPSPQLRYSEDDPVPLPSESTDVYVAPLTC 1133
 Qy 1140 SPOEYVYNOQDVRPQPSPREGLPAARPAAGATLERAKTLPQKNGVYKDVAFAGAVEN 1199
 Db 1134 SPOEYVYNOQDVRPQPSPREGLPAARPAAGATLERAKTLPQKNGVYKDVAFAGAVEN 1193
 Qy 1200 PEYITPQGAAPQHPAPSPAPDNLYWDDPDEGAPSTFKGFTENENYGLDY 1259
 Db 1194 PEYITPQGAAPQHPAPSPAPDNLYWDDPDEGAPSTFKGFTENENYGLDY 1253
 Qy 1260 PV 1261
 Db 1254 PV 1255

RESULT 5

AAW01111 standard; Protein: 1255 AA.

AC AAW01111;

DT 01-JAN-1997 (first entry)

DE HER-2/neu protein.

KM HER-2/neu; c-erbB1, p185; oncogene; tyrosine protein kinase;

KM breast cancer; ovary cancer; colon cancer; lung cancer;

KM prostate cancer; immunisation; tumour; vaccine; vector.

OS Homo sapiens.

PN MO9630514-A1.

PD 03-OCT-1996.

Location/Qualifiers
 676..1255
 /label= Intracellular domain
 /note= "claimed domain, useful for immunisation"

PF 28-MAR-1996; 96MO-US01689.
 XX
 PR 31-MAR-1995; 95US-0414417.
 XX
 PA (UNITW) UNITW WASHINGTON.
 XX
 PI Cheever MA, Disis ML;
 XX
 DR WPI; 1996-45361/45.
 XX
 DR N-PSDB; AAT40739.
 PT
 PT DNA encoding HER-2/neu polypeptide(s) - used for prevention or
 PT treatment of malignancies with which the HER-2/neu oncogene is
 PT associated
 PS
 PS Claim 2; Page 56-61; 71pp; English.
 CC
 CC Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
 CC the product of the HER-2/neu oncogene (see also AAT40739). The
 CC protein is over-expressed in various cancers, including breast,
 CC ovarian, colon, lung and prostate. The intracellular domain of the
 CC protein can be used to immunise an animal against a malignancy with
 CC which the oncogene is associated. The polypeptide can be produced
 CC in transformed host cells for use in immunisation. Alternatively,
 CC animal cells are transfected in vivo or ex vivo with a viral vector
 CC that directs expression of the polypeptide.
 CC
 SQ Sequence 1255 AA:
 Query Match 96.9%; Score 6641; DB 17; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
 Qy 1 MELALCEWGLLALPPGAASYCTGTDMLRLPASPETHLMDLRHLYGQCYVQGNL 60
 Db 1 MELALCEWGLLALPPGAASYCTGTDMLRLPASPETHLMDLRHLYGQCYVQGNL 60
 Qy 61 ELTYLPTNASLFLDIOEVQGYVLIANQVRYVPLQRLRYRGTLQFEDYVALVLDNG 120
 Db 61 ELTYLPTNASLFLDIOEVQGYVLIANQVRYVPLQRLRYRGTLQFEDYVALVLDNG 120
 Qy 121 DPLNNTPTGASPGELRELOSLTEILKGSVLIQRNPOLCYODTILMKDIFHNQOLA 180
 Db 121 DPLNNTPTGASPGELRELOSLTEILKGSVLIQRNPOLCYODTILMKDIFHNQOLA 180
 Qy 181 LTLIDTNSRACHPCSPCKGRCWGESSEDCQSLTRTVACGACARCKGRLPTDCHEQC 240
 Db 181 LTLIDTNSRACHPCSPCKGRCWGESSEDCQSLTRTVACGACARCKGRLPTDCHEQC 240
 Qy 241 AAGCTGPRGSDCLAHFNHSGICELHCPALVYNTDTFESMPNPEGRTTGASCVTACP 300
 Db 241 AAGCTGPRGSDCLAHFNHSGICELHCPALVYNTDTFESMPNPEGRTTGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPFLNQEVTAEDGTQRCCKSKPCARVCYGMQYIRANSKFIGIT 360
 Db 301 YNYLSTDVGSCTLVCPFLNQEVTAEDGTQRCCKSKPCARVCYGMQYIRANSKFIGIT 360
 Qy 361 ELFPAGCKKIFGSLAFLEPSFDGPASTATLQEOLOVFTLEITGYLYISAMPDPLP 420
 Db 361 ELFPAGCKKIFGSLAFLEPSFDGPASTATLQEOLOVFTLEITGYLYISAMPDPLP 420
 Qy 421 DLSVFQNLQVIRGRIILHNGAYSLLQGGIGISMLGRLSRLREGSGALIIHNTHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIILHNGAYSLLQGGIGISMLGRLSRLREGSGALIIHNTHLCFVHTV 480
 Qy 481 PMDQLFNPHQALILHTANRPEDCEVGGGLACHQICARGHCWGPPTQCVNCSQFLRQEC 540
 Db 481 PMDQLFNPHQALILHTANRPEDCEVGGGLACHQICARGHCWGPPTQCVNCSQFLRQEC 540
 Qy 541 VEECRVIGQLPREYVNAHCLPCHPECOPONGSVTCGPREDDQVCAAHKDPFCVARC 600
 Db 541 VEECRVIGQLPREYVNAHCLPCHPECOPONGSVTCGPREDDQVCAAHKDPFCVARC 600

QY 601 PSQVKEPDLSTMPIMKPFDEGACQPCINCTHSCVLDLDDKGCAPQASPLTSTVAVNG 660
 DB 601 PSQVKEPDLSTMPIMKPFDEGACQPCINCTHSCVLDLDDKGCAPQASPLTSTVAVNG 660
 QY 661 ILVVLGVGVGGLIKRQOKIRKRYTMRRLTETELVEPLTPSGAPNQAQVILKEFN 720
 DB 661 ILVVLGVGVGGLIKRQOKIRKRYTMRRLTETELVEPLTPSGAPNQAQVILKEFE 719
 QY 721 FTVSFWLVRVSVAS-HLETYKGIWIPDGENVKIPVALKLENTSPKANKEILDEAVY 779
 DB 720 -----LRKVKVLSGAGFTVYKGIWIPDGENVKIPVALKLENTSPKANKEILDEAVY 773
 QY 780 MAGVSPYVSRLLIGICTSTVQVLTQMLPQGLLDVRENRGLSGODLNMOMQIAKGM 839
 DB 774 MAGVSPYVSRLLIGICTSTVQVLTQMLPQGLLDVRENRGLSGODLNMOMQIAKGM 833
 QY 840 SYLEDVLRVLRVSVAS-HLETYKGIWIPDGENVKIPVALKLENTSPKANKEILDEAVY 899
 DB 834 SYLEDVLRVLRVSVAS-HLETYKGIWIPDGENVKIPVALKLENTSPKANKEILDEAVY 893
 QY 900 ILRRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVTM 959
 DB 894 ILRRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVTM 953
 QY 960 IMTKCMWIDSECRPRRELVSFSPMAADPQRFVVIQNEGLGASPLDSTFTYSLLEDD 1019
 DB 954 IMTKCMWIDSECRPRRELVSFSPMAADPQRFVVIQNEGLGASPLDSTFTYSLLEDD 1013
 QY 1020 MGDLVDAEYLVPQGGFCPPAPAGAGVHRRHSSSTRSGGDLTTLGLEPSEEPAPRS 1079
 DB 1014 MGDLVDAEYLVPQGGFCPPAPAGAGVHRRHSSSTRSGGDLTTLGLEPSEEPAPRS 1073
 QY 1080 PLAPSEGAGSDVDGDLGMAAKGLOSLPTHDSPLQRYSEDTVLPSETGYVAPLTC 1139
 DB 1074 PLAPSEGAGSDVDGDLGMAAKGLOSLPTHDSPLQRYSEDTVLPSETGYVAPLTC 1133
 QY 1140 SPQPEYVNDVAPQPPSPREGPLPAAPAGATLEPAKTLSPGKNGVVDVAFGAVEN 1199
 DB 1134 SPQPEYVNDVAPQPPSPREGPLPAAPAGATLEPAKTLSPGKNGVVDVAFGAVEN 1193
 QY 1200 PEVLTPQGGAPQPPSPAPFNDLYWDDQPPERGAPESTFKGTPTAENPEYGLDV 1259
 DB 1194 PEVLTPQGGAPQPPSPAPFNDLYWDDQPPERGAPESTFKGTPTAENPEYGLDV 1253
 QY 1260 PV 1261
 DB 1254 PV 1255
 RESULT 6
 AAM92406
 ID AAM92406 standard; Protein; 1255 AA.
 AC AAM92406;
 XX
 DT 21-APR-1999 (first entry)
 XX
 DE Human HER-2/neu oncogene protein.
 XX
 KM HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;
 XX malignancy; treatment; tumour.
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT Region 676..1255
 FT /note="region which elicits immune response"
 XX
 PN US5869445-A.
 XX
 PD 09-FEB-1999.
 XX
 PF 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.
 PR 17-MAR-1993; 93US-0033644.
 PR 12-AUG-1993; 93US-0106112.
 PR 31-MAR-1995; 95US-0414417.
 PA (UNIT) UNIT WASHINGTON.
 XX
 PI Cheever MA, Disis ML;
 XX
 DR WPI: 1999-152835/13.
 DR N-PSDB; AAX01912.
 XX
 PT Use of HER-2/neu polypeptides - for eliciting an immune response to
 PT an HER-2/neu associated malignancy, particularly for treating or
 PT preventing tumours
 PS
 CC Claim 3; Column 31-38; 26pp; English.
 CC
 CC This sequence represents the human HER-2/neu oncogene protein. A fragment
 CC of this protein is used in a method for eliciting or enhancing an immune
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
 CC B cells to produce an immune response to the HER-2/neu protein. The
 CC method can be used for immunisation against a malignancy in which the
 CC HER-2/neu oncogene is associated and in the treatment of an existing
 CC tumour, or to prevent tumour occurrence or recurrence.
 CC
 XX
 SQ Sequence 1255 AA;
 Query Match 96.9%; Score 6641; DB 20; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
 1 MELAALCRWGLLALLPFGAASVQVCTGDMKRLPASPETHLMLRLYQCCQVVGNTL 60
 1 MELAALCRWGLLALLPFGAASVQVCTGDMKRLPASPETHLMLRLYQCCQVVGNTL 60
 61 ELYTPVNASLSFLQDIQEVQGYVLIANOVQVPLQRLRVGTQLFEDVYALAVDNG 120
 61 ELYTPVNASLSFLQDIQEVQGYVLIANOVQVPLQRLRVGTQLFEDVYALAVDNG 120
 121 DPLNNTPTVTGASPEGLREQLRSITELIKGVALIQRNPOLCYODTIMKDIPIKNNOLA 180
 121 DPLNNTPTVTGASPEGLREQLRSITELIKGVALIQRNPOLCYODTIMKDIPIKNNOLA 180
 161 LFLIDNRSRACHPCSPWCKGSRCKWGESSEDCQSLTRTVCAAGCARKGRLPTDCHEQC 240
 161 LFLIDNRSRACHPCSPWCKGSRCKWGESSEDCQSLTRTVCAAGCARKGRLPTDCHEQC 240
 181 LFLIDNRSRACHPCSPWCKGSRCKWGESSEDCQSLTRTVCAAGCARKGRLPTDCHEQC 240
 241 AAGCTGPKGSDCLCLHFNHSGICEIHPALVTNTDFEEMPMDEGRYTGAACVACP 300
 241 AAGCTGPKGSDCLCLHFNHSGICEIHPALVTNTDFEEMPMDEGRYTGAACVACP 300
 301 YNYLSTDVGSCTLVCPHNDVETAEADGTQRCCKSKPCARVCYGLGMQYIRANSKFIGIT 360
 301 YNYLSTDVGSCTLVCPHNDVETAEADGTQRCCKSKPCARVCYGLGMQYIRANSKFIGIT 360
 361 ELYFAGCKKITGSLAFPLPESPDGDPASTALQEOQLVFTLEITGYLISAMPDPLP 420
 361 ELYFAGCKKITGSLAFPLPESPDGDPASTALQEOQLVFTLEITGYLISAMPDPLP 420
 421 DLVSFQNLQVIRGRILHNGAYSLTQGGISWMLGRSLRELGSGLALHNHHTHLCFYHTV 480
 421 DLVSFQNLQVIRGRILHNGAYSLTQGGISWMLGRSLRELGSGLALHNHHTHLCFYHTV 480
 481 FMDQLFRNPHQALHTANRPEDCEVGGGLACHQICARGHCGPPTQCVNCSQFLRQEC 540
 481 FMDQLFRNPHQALHTANRPEDCEVGGGLACHQICARGHCGPPTQCVNCSQFLRQEC 540
 541 VEECRVIGQLPREYVNAHCLCPGPECCOPONGSVTCGPEADQCVAAAHYKDPFCVAC 600
 541 VEECRVIGQLPREYVNAHCLCPGPECCOPONGSVTCGPEADQCVAAAHYKDPFCVAC 600

QY 601 PSQVDPDLSYMPIWKPFDEEGACOPCPINCTHSCVDLDDKCEPAEQRASPLTIVSAVVG 660
 DB 601 PSQVDPDLSYMPIWKPFDEEGACOPCPINCTHSCVDLDDKCEPAEQRASPLTIVSAVVG 660
 QY 661 ILLVVLGVVFGILIKRQCKIRKYTRRLLOETELVEPLTSGAMPNQAOMILKEFNN 720
 DB 661 ILLVVLGVVFGILIKRQCKIRKYTRRLLOETELVEPLTSGAMPNQAOMILKEFNN 720
 QY 721 FTYSFWLRVPSVAS -HLETVYKGIWIPDGENVKIPAVIVLENNSPKANKKILDEAVY 779
 DB 720 -----LRKVKVLQSGAFVTVYKGIWIPDGENVKIPAVIVLENNSPKANKKILDEAVY 773
 QY 780 MAGVSPYVRLDGLCTSTVOLVQLMPYGCILLDVHRENRGLSGODLLNMCQIAKGM 839
 DB 774 MAGVSPYVRLDGLCTSTVOLVQLMPYGCILLDVHRENRGLSGODLLNMCQIAKGM 833
 QY 840 SYLEDVLRHDLAARVLYKSPHVKITDPGLARLLDDETYHADGGKVPYIKMALES 869
 DB 834 SYLEDVLRHDLAARVLYKSPHVKITDPGLARLLDDETYHADGGKVPYIKMALES 863
 QY 900 ILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKGERLPQPICTIDVYM 959
 DB 894 ILRRRPTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKGERLPQPICTIDVYM 953
 QY 960 IMVKCMWIDSECRPRELVESESRMARDPQRFVYIIONEDLGASPLDSTFYRSLIEDDD 1019
 DB 954 IMVKCMWIDSECRPRELVESESRMARDPQRFVYIIONEDLGASPLDSTFYRSLIEDDD 1013
 QY 1020 MGDLVDAEEYLVPOQGFCEPDPAPAGAGMWHHRSSSTSGGGDTLGLPESEEARPS 1079
 DB 1014 MGDLVDAEEYLVPOQGFCEPDPAPAGAGMWHHRSSSTSGGGDTLGLPESEEARPS 1073
 QY 1080 PLASEGAGSDVEDDGLGMAAKGLQSLPTHDSPLQRYSEDPVLPSESTDGYAPLTC 1139
 DB 1074 PLASEGAGSDVEDDGLGMAAKGLQSLPTHDSPLQRYSEDPVLPSESTDGYAPLTC 1133
 QY 1140 SPOEYVNOQDVAPQPSREGPLPARPAGATLERKTLSPKNGVAKDVFAGAVEN 1199
 DB 1134 SPOEYVNOQDVAPQPSREGPLPARPAGATLERKTLSPKNGVAKDVFAGAVEN 1193
 QY 1200 PEYLTPOGGAPOHPHPAPSPAFDNLVYWDODPERGAPSTFKGPTAENPEYGLDY 1259
 DB 1194 PEYLTPOGGAPOHPHPAPSPAFDNLVYWDODPERGAPSTFKGPTAENPEYGLDY 1253
 QY 1260 PV 1261
 DB 1254 PV 1255

RESULT 7
 AAB21198
 ID AAB21198 standard; protein: 1255 AA.

AC AAB21198;
 DT 12-JAN-2001 (first entry)
 XX
 DE Human HER-2/neu protein.
 XX
 KW Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200044899-A1.
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000MO-US02164.
 XX
 PR 29-JAN-1999; 99US-0117976.

PA (CORI-) CORIXA CORP.
 PA (SMK) SMITHKLINE BEECHAM.
 XX
 PI Cheever MA, Cheysen D;
 XX
 DR WPI, 2000-505976/45.
 DR N-PSDB; AAA89736.
 XX
 PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX
 PS Claim 52; Fig 7; 128pp; English.
 XX
 CC The present sequence is the human HER-2/neu protein. It is a member of
 CC the tyrosine kinase family of receptor-like glycoproteins and shows
 CC homology to the epidermal growth factor receptor (EGFR). It probably
 CC plays a part in cell growth and/or differentiation. The HER-2/neu
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against
 CC these neoplasias.
 CC
 SQ Sequence 1255 AA;
 Query Match 96.9%; Score 6641; DB 21; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
 QY 1 MELALCFWGLLALLPFGAASVYCTGTMKLRPASPTHLDMLRHLVQGCQVVGNTL 60
 DB 1 MELALCFWGLLALLPFGAASVYCTGTMKLRPASPTHLDMLRHLVQGCQVVGNTL 60
 QY 61 ELTYPTNASSFLDDIOBVQGVLIANNOVQVPLQRLRYRGTLQLEEDNYALAVDNG 120
 DB 61 ELTYPTNASSFLDDIOBVQGVLIANNOVQVPLQRLRYRGTLQLEEDNYALAVDNG 120
 QY 121 DLNNTPTATGASPGGLRELORSLETILKGVLIQRNPOLCYQDTILMKDIFHGNOLA 180
 DB 121 DLNNTPTATGASPGGLRELORSLETILKGVLIQRNPOLCYQDTILMKDIFHGNOLA 180
 QY 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLETDCHEQC 240
 DB 181 LTLIDNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLETDCHEQC 240
 QY 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDFEEMPNREGYTGASCVTACP 300
 DB 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDFEEMPNREGYTGASCVTACP 300
 QY 301 YNYLSTDVSGCTLVCPFLHNOEYTAEDGTQCEKCSKPCARVYCGLGKQYIKANSKFIGIT 360
 DB 301 YNYLSTDVSGCTLVCPFLHNOEYTAEDGTQCEKCSKPCARVYCGLGKQYIKANSKFIGIT 360
 QY 361 ELFEAGCKKIFGSLAFLEPSFDGDPASNTAPLOPQLOVFFTLBEITGYLIISAMPDLP 420
 DB 361 IQEFAQCKKIFGSLAFLEPSFDGDPASNTAPLOPQLOVFFTLBEITGYLIISAMPDLP 420
 QY 421 DLSTVFQNIQVIRGLIHNGAVSLTQIGISIMLGIRSLRELGSGALTLHHNTHLCFVATV 480
 DB 421 DLSTVFQNIQVIRGLIHNGAVSLTQIGISIMLGIRSLRELGSGALTLHHNTHLCFVATV 480
 QY 481 FWDOLFRRNHQALLHTANRPEDECVGEGIALCHQLCARGHCWGPPTQCVCNSQFLRGEC 540
 DB 481 FWDOLFRRNHQALLHTANRPEDECVGEGIALCHQLCARGHCWGPPTQCVCNSQFLRGEC 540
 QY 541 VEEGVTLQGLPREYVNAHCHLPCHPECCPQNGSVTCFPEEDQCACAHYKDPFCVARC 600
 DB 541 VEEGVTLQGLPREYVNAHCHLPCHPECCPQNGSVTCFPEEDQCACAHYKDPFCVARC 600
 QY 601 PSQVDPDLSYMPIWKPFDEEGACOPCPINCTHSCVDLDDKCEPAEQRASPLTIVSAVVG 660

```

Db      601 PSQVPLSLTPIPKFDEBACOPCINCTHSCVDLDDGCPPEQASPLTSTISAVG 660
QY      661 ILVVVLGVVFGILLIKRQOKIRKXTRRLLOETELVEPLTPSGAMNQAKMILKEFNN 720
Db      661 ILVVVLGVVFGILLIKRQOKIRKXTRRLLOETELVEPLTPSGAMNQAKMILKEFNN 719
QY      721 FTVSFMRVPRKVSAS-HLETVYKGIWIPDGENVKIPALVYLENTPSPKANKILDEAY 779
Db      720 -----LRKVKVSGSGAGTVYKGIWIPDGENVKIPALVYLENTPSPKANKILDEAY 773
QY      780 MAGVSPYVSRLIGICLTSTVQVLTQMPYGCLLDHYRENGRLGSDLLNMCQIAKGM 839
Db      774 MAGVSPYVSRLIGICLTSTVQVLTQMPYGCLLDHYRENGRLGSDLLNMCQIAKGM 833
QY      840 SYLEDVTLVRDLAARVLYKSPNHVKITDFGLARLIDIDETEVHADGKVPKIMMALES 899
Db      834 SYLEDVTLVRDLAARVLYKSPNHVKITDFGLARLIDIDETEVHADGKVPKIMMALES 893
QY      900 ILRRFTHSDVMSYGVTVWELMTFGAKPYDGI PAPEIPDLLEKGERLPQPICTIDVYM 959
Db      894 ILRRFTHSDVMSYGVTVWELMTFGAKPYDGI PAPEIPDLLEKGERLPQPICTIDVYM 953
QY      960 IMVKCMWIDSECRPRFRELVSFESRMARDPQRFVYIQNEDLGPASPLDSTFYRLDEDD 1019
Db      954 IMVKCMWIDSECRPRFRELVSFESRMARDPQRFVYIQNEDLGPASPLDSTFYRLDEDD 1013
QY      1020 MGDLVAEEYLVPOQGFCEPDPAAGAGMTHHRSSSTSSGGGDLTLGIEPEEAPRS 1079
Db      1014 MGDLVAEEYLVPOQGFCEPDPAAGAGMTHHRSSSTSSGGGDLTLGIEPEEAPRS 1073
QY      1080 PLAPSEGAGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDTVPLPSETDYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDTVPLPSETDYVAPLTC 1133
QY      1140 SPPEYVNOVDVAPQPPSPREGPLPAARPAAGATLERPKTLSPGKNVVDVAFGAVEN 1199
Db      1134 SPPEYVNOVDVAPQPPSPREGPLPAARPAAGATLERPKTLSPGKNVVDVAFGAVEN 1193
QY      1200 PEYLTPOGGAAPQPPSPAFDNLVYWDOPPEPRGAPPESTFKGTPTAENPEYGLDV 1259
Db      1194 PEYLTPOGGAAPQPPSPAFDNLVYWDOPPEPRGAPPESTFKGTPTAENPEYGLDV 1253
QY      1260 PV 1261
Db      1254 PV 1255

RESULT 8
AA84780
ID      AAY84780 standard; Protein; 1255 AA.
XX
AC      AAY84780;
XX
DT      08-AUG-2000 (first entry)
XX
DE      Amino acid sequence of the SPLICE erbb-2 receptor protein.
XX
KW      SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW      tumor cell proliferation; tissue degeneration; arthropathy;
KW      bone resorption; inflammatory disease; degenerative disorder;
XX
OS      Homo sapiens.
XX
PN      WO200020579-A1.
XX
PD      13-APR-2000.
XX
PF      01-OCT-1999; 99WO-CA00912.
XX
PR      02-OCT-1998; 98US-0165192.
XX

```

```

PA      (UIMC-) UNITV MCMMASTER.
PI      Muller WJ, Siegel PM,
XX      WPI: 2000-303768/26.
DR      N-PSDB; AAA14612.
XX
PT      Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
XX      erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX      Claim 3; Fig 2; 60pp; English.
XX
CC      The present sequence represents a SPLICE erbb-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erbb-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
XX
SQ      Sequence 1255 AA;
XX
Query Match 96.9%; Score 6641; DB 21; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
QY      1 MELALCWGILLALPPGAASVQCTGTMKLRIPASPEHLDMLRLHYOGCQVQGNL 60
Db      1 MELALCWGILLALPPGAASVQCTGTMKLRIPASPEHLDMLRLHYOGCQVQGNL 60
QY      61 EETVPTVASISFLDIDIEVQGYVLIANOVROYPLQRLTVRGQLPEEDYVALAVLDNG 120
Db      61 EETVPTVASISFLDIDIEVQGYVLIANOVROYPLQRLTVRGQLPEEDYVALAVLDNG 120
QY      121 DPLNNTPTVGTASPGGLRELQRLSTELIKGVLIORNPOLCYODTILMKDIFHNQOLA 180
Db      121 DPLNNTPTVGTASPGGLRELQRLSTELIKGVLIORNPOLCYODTILMKDIFHNQOLA 180
QY      181 LTLIDTNSRACHGSPKCKSRCKWGESSECCSLTRVCGAGCARCKGPIPTDCHEQC 240
Db      181 LTLIDTNSRACHGSPKCKSRCKWGESSECCSLTRVCGAGCARCKGPIPTDCHEQC 240
QY      241 AAGCTGPGSDCLACLFHNSGICELHPALVYNTDPEFSPMPDEGRYTGCASCVTACP 300
Db      241 AAGCTGPGSDCLACLFHNSGICELHPALVYNTDPEFSPMPDEGRYTGCASCVTACP 300
QY      301 YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVYCYGLQMOYIKANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVYCYGLQMOYIKANSKFIGIT 360
QY      361 ELEFAGCKKIFGSLAFIPESFDPASNTAPLOEOUQVETTELITGYLTISAMPDPLP 420
Db      361 ELEFAGCKKIFGSLAFIPESFDPASNTAPLOEOUQVETTELITGYLTISAMPDPLP 420
QY      421 DLSVFQMLQVIRGRILHNQAVSLTLQIGISWLGRLRLRELSGLALIHNTHTLCFYHTV 480
Db      421 DLSVFQMLQVIRGRILHNQAVSLTLQIGISWLGRLRLRELSGLALIHNTHTLCFYHTV 480
QY      481 PMDQLFRNPHQALHTANRPEDDEVGGLACHQICANGHCKMGPPTCCVNCOSQFLRQEC 540
Db      481 PMDQLFRNPHQALHTANRPEDDEVGGLACHQICANGHCKMGPPTCCVNCOSQFLRQEC 540
QY      541 VEECRVQLPREYVNAARHCLPCHPECOPOGNGSYTCGPEADQCVACAHYKDPFFCYARC 600
Db      541 VEECRVQLPREYVNAARHCLPCHPECOPOGNGSYTCGPEADQCVACAHYKDPFFCYARC 600
QY      601 PSQVPLSLTPIPKFDEBACOPCINCTHSCVDLDDGCPPEQASPLTSTISAVG 660

```

```

Db 601 PSQVPSDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTISIISAVG 660
Qy 661 ILLVVVGVVFGILLIKRQOKIRKXTMRLLLOETELVEPLTPSGAMPQAOQMRLLKEENN 720
Db 661 ILLVVVGVVFGILLIKRQOKIRKXTMRLLLOETELVEPLTPSGAMPQAOQMRLLKEETE- 719
Qy 721 FTVSEMLRVPKVSAS-NLETVYKGIWIPDGENVKIPVAIKVLRNENTSEKANKELIDEAIV 779
Db 720 -----LKKVVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNENTSEKANKELIDEAIV 773
Qy 780 MAGVSPVYSRLILGICLTSTVQVLTQMLPFGCLLDHRENRGRIGSQDLNMCQIAKGM 839
Db 774 MAGVSPVYSRLILGICLTSTVQVLTQMLPFGCLLDHRENRGRIGSQDLNMCQIAKGM 833
Qy 840 SYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLIDETETHADGKVPIMKMALES 899
Db 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLIDETETHADGKVPIMKMALES 893
Qy 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYM 959
Db 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYM 953
Qy 960 IMVKCMWIDSECRPFRELVSEFSMARDPQRFVYIQNEDLGASPLDSTFYSRLLEDD 1013
Db 954 IMVKCMWIDSECRPFRELVSEFSMARDPQRFVYIQNEDLGASPLDSTFYSRLLEDD 1013
Qy 1020 MGDVDAEYLVPOQGFPCPDPAAGAGMVRHRRSSSTRSGGDLTLGLEPSEEAARS 1079
Db 1014 MGDVDAEYLVPOQGFPCPDPAAGAGMVRHRRSSSTRSGGDLTLGLEPSEEAARS 1073
Qy 1080 PLARSEGAGSDVFGDGLMGAAKGLQSLPTHDSPLQKYSDDPVPPLSEVDGYVAPLTC 1139
Db 1074 PLARSEGAGSDVFGDGLMGAAKGLQSLPTHDSPLQKYSDDPVPPLSEVDGYVAPLTC 1133
Qy 1140 SPOPEYVQNPDPVRPQPPSPREGPLPAAPAGATLERAKTLPKNGKGVKQVFAFGAVEN 1199
Db 1134 SPOPEYVQNPDPVRPQPPSPREGPLPAAPAGATLERAKTLPKNGKGVKQVFAFGAVEN 1193
Qy 1200 PEYLTPOGGAAPQHPHPAPAFADNLYYMDQDPPERGAPSTFKGTPTAENPEYLGLDV 1259
Db 1194 PEYLTPOGGAAPQHPHPAPAFADNLYYMDQDPPERGAPSTFKGTPTAENPEYLGLDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
XX oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN MO200153463-A2.
XX
PD 26-JUL-2001.
XX
PE 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX

```

```

PI Cheever MA, Hand-Zimmermann S;
XX
DR WPI: 2001-476112/51.
XX N-PSDB; AAB23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
XX
Query Match 96.9%; Score 6641; DB 22; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
Qy 1 METALCRMGILLALPPGASSTQVCTGDMKRLPASPETHLDMRLHYOGQVQGNL 60
Db 1 METALCRMGILLALPPGASSTQVCTGDMKRLPASPETHLDMRLHYOGQVQGNL 60
Qy 61 ELTYLPTNASSLSLQDIOEVQGVLLIAHQVQVPLQRLIRVGTQLFEDNALVALDNG 120
Db 61 ELTYLPTNASSLSLQDIOEVQGVLLIAHQVQVPLQRLIRVGTQLFEDNALVALDNG 120
Qy 121 DPLNNTPTTGAAPGGLRELOLRSLTEILKGGVLIQRNQLCYQDITILMKDIFHKNNOLA 180
Db 121 DPLNNTPTTGAAPGGLRELOLRSLTEILKGGVLIQRNQLCYQDITILMKDIFHKNNOLA 180
Qy 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGCARKGSLPTDCHEQC 240
Db 181 LTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGCARKGSLPTDCHEQC 240
Qy 241 AACCTGPKSDDLACLHFNHSGICEHCAVALTYNTDTRESNPNBGRRTFFGASCTTAC 300
Db 241 AACCTGPKSDDLACLHFNHSGICEHCAVALTYNTDTRESNPNBGRRTFFGASCTTAC 300
Qy 301 YNYLSTDVSGCTVCPHNOEVTAEEDTORCEKSPKARVCYGLCMOYIKANSKFIGIT 360
Db 301 YNYLSTDVSGCTVCPHNOEVTAEEDTORCEKSPKARVCYGLCMOYIKANSKFIGIT 360
Qy 361 ELEFPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEITGLYISAWPDSL 420
Db 361 ELEFPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEITGLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILQGLISMLGARSJRELGSGLAIHHNHTLGVHIV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILQGLISMLGARSJRELGSGLAIHHNHTLGVHIV 480
Qy 481 PMDQFRNPHQALLHTANRPEDECVGEGJACHQLCARGHCWGGPTQCNCSQFLRGQGC 540
Db 481 PMDQFRNPHQALLHTANRPEDECVGEGJACHQLCARGHCWGGPTQCNCSQFLRGQGC 540
Qy 541 VEECRVLOGLPREYVNAARCLCHPECCQONSVCYTFGBEAQCVACATYKDPFVAVAC 600
Db 541 VEECRVLOGLPREYVNAARCLCHPECCQONSVCYTFGBEAQCVACATYKDPFVAVAC 600
Qy 601 PSQVPSDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTISIISAVG 660
Db 601 PSQVPSDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKCGPAEQRASPLTISIISAVG 660
Qy 661 ILLVVVGVVFGILLIKRQOKIRKXTMRLLLOETELVEPLTPSGAMPQAOQMRLLKEENN 720

```

Db 661 ILVVVLGVVFGILIKRRQOKIKYTKRRLLOETELVEPLTPSGAMPNQAQNFILKETE 719
 QY 721 FTVSFWLVRPKVSAS-HLETVYKGIWIPDGENYKIPVAIKVRENTSPKANKEILDEAVY 779
 Db 720 -----LRKVKVLGSGAFGTGYKGIWIPDGENYKIPVAIKVRENTSPKANKEILDEAVY 773
 QY 780 MAGVSPYVRLIGLICLTSTVCLVTOUMPYGCILDVHRENRKGLSGSDILNMOMAKKM 839
 Db 774 MAGVSPYVRLIGLICLTSTVCLVTOUMPYGCILDVHRENRKGLSGSDILNMOMAKKM 833
 QY 840 SYLEDVRLVHRLAARVNLVKSNNHVKITDFGLARLLDIDETEYHAHGGKVPKIMMALES 899
 Db 834 SYLEDVRLVHRLAARVNLVKSNNHVKITDFGLARLLDIDETEYHAHGGKVPKIMMALES 893
 QY 900 ILRRRFTHSDVMSYGVTTWELMTFFGAKPYDGIPEAEIPDLLEKGRLLPQPICTIDVYM 959
 Db 894 ILRRRFTHSDVMSYGVTTWELMTFFGAKPYDGIPEAEIPDLLEKGRLLPQPICTIDVYM 953
 QY 960 IMKCMWIDSECRPRRELVESEFSRMAPRQRFVVIQNEDLGASPDLSTFYRSLLEDD 1019
 Db 954 IMKCMWIDSECRPRRELVESEFSRMAPRQRFVVIQNEDLGASPDLSTFYRSLLEDD 1013
 QY 1020 MGDLVDAEYLVPOQGFPCPDPAFGAGVNHRRSSSTRSGGDLTLGLPSESEAPRS 1079
 Db 1014 MGDLVDAEYLVPOQGFPCPDPAFGAGVNHRRSSSTRSGGDLTLGLPSESEAPRS 1073
 QY 1080 PLAPSEAGSDVDGDLGGAAGKLSLPHDPSLQKRSSEPTVPLPSETGYAPLTC 1139
 Db 1074 PLAPSEAGSDVDGDLGGAAGKLSLPHDPSLQKRSSEPTVPLPSETGYAPLTC 1133
 QY 1140 SPQPEYVNPQVAPRPPSPREGEPLPAARAGATLEBAKTLSPGKNGVNDVAFAGAVEN 1199
 Db 1134 SPQPEYVNPQVAPRPPSPREGEPLPAARAGATLEBAKTLSPGKNGVNDVAFAGAVEN 1193
 QY 1200 PELTTPQGAAPQHPHPAPSPAFDNLVYWDQPPERGAAPSTFKCTPTAENPEYGLDV 1259
 Db 1194 PELTTPQGAAPQHPHPAPSPAFDNLVYWDQPPERGAAPSTFKCTPTAENPEYGLDV 1253
 QY 1260 PV 1261
 Db 1254 PV 1255

RESULT 10
 AAG88267
 ID AAG88267 standard; Protein; 1255 AA.
 AC AAG88267;
 XX
 DT 11-SEP-2001 (first entry)
 XX
 DE HER2/neu amino acid sequence.
 XX
 KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
 KW immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
 KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
 XX
 OS Homo sapiens.
 PN WO200141787-A1.
 PD 14-JUN-2001.
 PF 11-DEC-2000; 2000MO-US33591.
 PR 10-DEC-1999; 99US-0458299.
 PA (EPI-M-) EPIMUNE INC.
 PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
 DR WPI, 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
 PT cellular immune responses for the prevention and treatment of cancer
 XX
 PS Disclosure, Page 15; 199pp; English.

CC The present invention describes isolated prepared HER2/neu epitopes (I).
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
 CC culture in vitro and binds to a complex of an epitope (II), bound to a
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
 CC and a second epitope and the peptide is less than 50 contiguous amino
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
 CC and immunostimulant activities, and can be used in vaccines (I), (II)
 CC and (III) are useful for inducing cellular immune responses for the
 CC prevention and treatment of cancer. (I) and (II) are useful for
 CC monitoring or evaluating an immune response to a tumour-associated
 CC antigen when incubated with a T lymphocyte sample from a patient and
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
 CC based vaccines mean that immunosuppressive epitopes that may be present
 CC in whole antigens may be avoided. Selected epitopes may be combined to
 CC enhance immunogenicity. The possible pathological side effects caused by
 CC infectious agents or whole protein antigen is eliminated. The vaccine
 CC provides the ability to direct and focus an immune response to multiple
 CC selected antigens from the same pathogen. Epitope-based anti-tumour
 CC vaccines provides the opportunity to combine epitopes derived from
 CC multiple tumour-associated molecules addressing the problem of tumour-
 CC tumour variability and reducing the likelihood of tumour escape due to
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
 CC the exemplification of the present invention.

Sequence 1255 AA;

Query Match 96.9%; Score 6641; DB 22; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

QY 1 MELALCRWGLLALLPFGAASVCTGTDKMLPASPETHLMDLHLYOGCOVQGNL 60
 Db 1 MELALCRWGLLALLPFGAASVCTGTDKMLPASPETHLMDLHLYOGCOVQGNL 60
 QY 61 ELTYLPTNASLSPFQDIQEVGYVLIHNOVQVPLQRLRVRGTOLEFEDYALAVLDNG 120
 Db 61 ELTYLPTNASLSPFQDIQEVGYVLIHNOVQVPLQRLRVRGTOLEFEDYALAVLDNG 120
 QY 121 DPLNNTTPVTGASPGGARELOLRSLTELKGVLIQBNPOLCYGDTLLMKDIFHKNNOLA 180
 Db 121 DPLNNTTPVTGASPGGARELOLRSLTELKGVLIQBNPOLCYGDTLLMKDIFHKNNOLA 180
 QY 121 DPLNNTTPVTGASPGGARELOLRSLTELKGVLIQBNPOLCYGDTLLMKDIFHKNNOLA 180
 Db 121 DPLNNTTPVTGASPGGARELOLRSLTELKGVLIQBNPOLCYGDTLLMKDIFHKNNOLA 180
 QY 181 LTLIDNRSRACHPCSPCKGSRGSGSSSEDQSLRTVYAGGACGKGPLPTDCHEOC 240
 Db 181 LTLIDNRSRACHPCSPCKGSRGSGSSSEDQSLRTVYAGGACGKGPLPTDCHEOC 240
 QY 241 AAGCTGKHSDCALCLFPHNSGICEPLVYNTPTFSMPRPBERYTPGASCVTACP 300
 Db 241 AAGCTGKHSDCALCLFPHNSGICEPLVYNTPTFSMPRPBERYTPGASCVTACP 300
 QY 241 AAGCTGKHSDCALCLFPHNSGICEPLVYNTPTFSMPRPBERYTPGASCVTACP 300
 Db 241 AAGCTGKHSDCALCLFPHNSGICEPLVYNTPTFSMPRPBERYTPGASCVTACP 300
 QY 301 YNYLSTDVGSCTIVCPAHNOEVAEDTQRCCKSKRCARVCYGLGQYIKANSKFIGIT 360
 Db 301 YNYLSTDVGSCTIVCPAHNOEVAEDTQRCCKSKRCARVCYGLGQYIKANSKFIGIT 360
 QY 361 ELFEAGCKTIFGSLAFIPESFDGDPASNTAPLOBOLOVETLEITGYLYISAMPDLP 420
 Db 361 IQEFAAGCKTIFGSLAFIPESFDGDPASNTAPLOBOLOVETLEITGYLYISAMPDLP 420
 QY 421 DLSVFOQLQVIRGRILHNGAYSLTLOGLISWLGSLRSLRSGGLALIHNTLCPHTY 480
 Db 421 DLSVFOQLQVIRGRILHNGAYSLTLOGLISWLGSLRSLRSGGLALIHNTLCPHTY 480
 QY 481 PMDOLFPHFOALLHTRNPEDEVCBGLACHQLCARGHCWGPPTQCVNCSQPLRGQEC 540
 Db 481 PMDOLFPHFOALLHTRNPEDEVCBGLACHQLCARGHCWGPPTQCVNCSQPLRGQEC 540

Db 481 PMDOLFRRPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPRPTQCVNCSOFLRGQEC 540
 Qy 541 VEEGRVYQGLPREYVNAARHCLPCHPECOFONGSVTCGFPADOCVACAHYKDPFCVARC 600
 Db 541 VEEGRVYQGLPREYVNAARHCLPCHPECOFONGSVTCGFPADOCVACAHYKDPFCVARC 600
 Qy 601 PSYVYKPLSYMPWPKPFDEEGACOPCPCINCHSCVDLDKGCAPACORASPLTSIVAVVG 660
 Db 601 PSYVYKPLSYMPWPKPFDEEGACOPCPCINCHSCVDLDKGCAPACORASPLTSIVAVVG 660
 Qy 661 ILVAVYLVGVFGLIKRRROOKIRKYTRMRILQETELVEPLTPSGAMNOMQMLKEFN 720
 Db 661 ILVAVYLVGVFGLIKRRROOKIRKYTRMRILQETELVEPLTPSGAMNOMQMLKEFN 719
 Qy 721 FTVSFWLVRPKVSAS-HLETYVKGIMV-PDGENVKIPVAIKVLEBNTSPKANKELDEAYV 779
 Db 720 -----LRKYVLSGSAFGTYVKGIMV-PDGENVKIPVAIKVLEBNTSPKANKELDEAYV 773
 Qy 780 MAGVGSFYVSRILGICITSTVQVLTQMLPYGCLLDHVRNKGRLGSDLLNMCQIAKGM 839
 Db 774 MAGVGSFYVSRILGICITSTVQVLTQMLPYGCLLDHVRNKGRLGSDLLNMCQIAKGM 833
 Qy 840 SYLEDVRLVHRDLAARVLYKSPVHVKITDGLARLLIDETEVHADGKVPFKMALES 899
 Db 834 SYLEDVRLVHRDLAARVLYKSPVHVKITDGLARLLIDETEVHADGKVPFKMALES 893
 Qy 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI-PAREIPDLLEKGERLPQPPICITDIVM 959
 Db 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI-PAREIPDLLEKGERLPQPPICITDIVM 953
 Qy 960 IMVKCMIMIDSECRFRFELVSEFSRMAKDPORFVIONEDLGPASPLDSTFYRSLLEDDO 1019
 Db 954 IMVKCMIMIDSECRFRFELVSEFSRMAKDPORFVIONEDLGPASPLDSTFYRSLLEDDO 1013
 Qy 1020 MGDVDAEELVLPQOGFCFDPAPAGAGMWHRRSSSTRSGGDLTLGLPESSEBAPRS 1079
 Db 1014 MGDVDAEELVLPQOGFCFDPAPAGAGMWHRRSSSTRSGGDLTLGLPESSEBAPRS 1073
 Qy 1080 PLAPSEGASDVEPGDGLGMAAKGLOSLPTHDPSPLOVSEDDTVLPSTDDVYAPLTC 1139
 Db 1074 PLAPSEGASDVEPGDGLGMAAKGLOSLPTHDPSPLOVSEDDTVLPSTDDVYAPLTC 1133
 Qy 1140 SPOEYVNOVDVSPQPSPREGRLPAARPAATLERAKTILSPKNGVYKDVFAFGAVEN 1199
 Db 1134 SPOEYVNOVDVSPQPSPREGRLPAARPAATLERAKTILSPKNGVYKDVFAFGAVEN 1193
 Qy 1200 PEYLTPOGGAAPQHPAPASPAFDNLYWDQDPPERGAPOSTFKGTPTANPEYVGLDV 1259
 Db 1194 PEYLTPOGGAAPQHPAPASPAFDNLYWDQDPPERGAPOSTFKGTPTANPEYVGLDV 1253
 Qy 1260 PV 1261
 Db 1254 PV 1255
 RESULT 11
 AAE24067
 ID AAE24067 standard; Protein: 1255 AA.
 XX AAE24067;
 DT 23-SEP-2002 (first entry)
 XX Human Her-2 protein.
 XX Human Her-2 protein.
 XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
 KM hyperproliferative disorder; prophylaxis; inflammation; antisense;
 KM tumour; gene therapy; phosphorocinicate backbone.
 XX Homo sapiens.
 OS Homo sapiens.
 PN MO200222636-A1.

PD 21-MAR-2002.
 XX 12-SEP-2001; 2001MO-US28572.
 XX 15-SEP-2000; 2000US-0663834.
 PR (ISIS-) ISIS PHARM INC.
 PA Bennett CF, Cowselet LM;
 PI MPI: 2002-471192/50.
 DR N-PSDB; AAD38904.
 XX Novel antisense oligonucleotide which modulates the expression of Human
 PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
 PT inflammation or to prevent infection in humans -
 XX Example 13; Page 95-107; 116pp; English.
 PS The invention relates to antisense compounds targeted to a nucleic
 CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
 CC that specifically hybridizes with and inhibits the expression of Her2.
 CC Antisense compounds of the invention are used for treating disorders or
 CC conditions associated with Her2 such as hyperproliferative disorders
 CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
 CC neural or cardiac cancer. They are also useful prophylactically e.g.
 CC to prevent or delay infection, inflammation and tumour formation. The
 CC invention is also used in gene therapy. The present sequence is human
 CC Her-2 protein.
 XX
 SQ Sequence 1255 AA;
 Query Match 96.9%; Score 6641; DB 23; Length 1255;
 Best Local Similarity 97.3%; Pred. No. 0;
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
 Qy 1 MELAALCWGLLALLPGASTQVCTGCTMKRLPASPETHLWLRILYQCCQVQGNL 60
 Db 1 MELAALCWGLLALLPGASTQVCTGCTMKRLPASPETHLWLRILYQCCQVQGNL 60
 Qy 61 ELTYLPTNASLFLDIOEVQYVLIANOVQVPLQSLRIVRGTLQFEDNYVALAVLNG 120
 Db 61 ELTYLPTNASLFLDIOEVQYVLIANOVQVPLQSLRIVRGTLQFEDNYVALAVLNG 120
 Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTELKGVLIQNPOLCYQDTILMKOIFKKNOLA 180
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTELKGVLIQNPOLCYQDTILMKOIFKKNOLA 180
 Qy 181 LTLIDTNSRACHPSPCKSGRSGESSEDCQSITRTVCAAGGCRGKPLPTDCHEOC 240
 Db 181 LTLIDTNSRACHPSPCKSGRSGESSEDCQSITRTVCAAGGCRGKPLPTDCHEOC 240
 Qy 241 AAGCTGPGHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTAC 300
 Db 241 AAGCTGPGHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTAC 300
 Qy 301 YNYLSTDVSGCTIVCPILNCEVTAEDGRCCKSPCARCYGUGWYIANSKFIGIT 360
 Db 301 YNYLSTDVSGCTIVCPILNCEVTAEDGRCCKSPCARCYGUGWYIANSKFIGIT 360
 Qy 361 ELIEFAGCKKIFGSLAFIPESFDGPASNTAFLQEPOLQVFETLEITGYLISAMPDLP 420
 Db 361 IOEFAGCKKIFGSLAFIPESFDGPASNTAFLQEPOLQVFETLEITGYLISAMPDLP 420
 Qy 421 DLSVQONQVIRGRILNHGAYSLTQIGISWLGRLSRLREGSGALHNNTHLCFVHTV 480
 Db 421 DLSVQONQVIRGRILNHGAYSLTQIGISWLGRLSRLREGSGALHNNTHLCFVHTV 480
 Qy 481 PMDOLFRRPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPRPTQCVNCSOFLRGQEC 540
 Db 481 PMDOLFRRPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPRPTQCVNCSOFLRGQEC 540
 Qy 541 VEEGRVYQGLPREYVNAARHCLPCHPECOFONGSVTCGFPADOCVACAHYKDPFCVARC 600


```

Db      541 VEECRVLYQGLPRYVNAHRLCHPEQPNQSVTFGPADQCVCAHYKPPFCVAC 600
Qy      601 PSQVDPDLSTYMPIMKFPDEBACQPCPINCTHSVDLDKCAEQRASPLTSIVAVG 660
Db      601 PSQVDPDLSTYMPIMKFPDEBACQPCPINCTHSVDLDKCAEQRASPLTSIVAVG 660
Qy      661 ILVVVLGVFGILIKRQCKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKEFN 720
Db      661 ILVVVLGVFGILIKRQCKIRKYMRLLOETELVEPLTPSGAMPNOQMRILKEFN 719
Qy      721 FTVSFMLRVPKVAS-HLETYKGIWIPDGENKIPVAIKVRENSPANKELIDEAV 779
Db      720 -----LRKVKVLGSAGFTVYKIMIPDGENKIPVAIKVRENSPANKELIDEAV 773
Qy      780 MAGVSPYVSRLLIGICTSTVQVLTQMPYGLLDHVRENRGLSGODLLNMCQIAKGM 839
Db      774 MAGVSPYVSRLLIGICTSTVQVLTQMPYGLLDHVRENRGLSGODLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARVLVKSNNHVKITDFGLARLLDIDETEHADGKVPIMKMALES 899
Db      834 SYLEDVRLVHRDLAARVLVKSNNHVKITDFGLARLLDIDETEHADGKVPIMKMALES 893
Qy      900 ILRRFTQSDVMSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVM 959
Db      894 ILRRFTQSDVMSYGVYTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVM 953
Qy      960 IMKCMWIDECRRRELSESRMARDQRVVITQNEDLGASPLDSTFTSRLLIEDD 1019
Db      954 IMKCMWIDECRRRELSESRMARDQRVVITQNEDLGASPLDSTFTSRLLIEDD 1013
Qy      1020 MGLDVAEEYLVQQGFCFCDPAPAGAGMWHHRSSSTRSGGGLTLGLEPSEBAPRS 1079
Db      1014 MGLDVAEEYLVQQGFCFCDPAPAGAGMWHHRSSSTRSGGGLTLGLEPSEBAPRS 1073
Qy      1080 PLAPSEGAGSDVFDGDLGMAKAGLSLPTHPDSPQRYSEDTVPJPSETDGYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGDLGMAKAGLSLPTHPDSPQRYSEDTVPJPSETDGYVAPLTC 1133
Qy      1140 SPQEVYNQDVNPPSPREBGLPARAPAGATLEBAKTLSPGKGNVMDVPAFGAVEN 1199
Db      1134 SPQEVYNQDVNPPSPREBGLPARAPAGATLEBAKTLSPGKGNVMDVPAFGAVEN 1193
Qy      1200 PEYLTPQGAAPQPHPPAPAFSPAFDNLVYWDQDPPEBGAPEPSTFKCTPTAENPEYLGIDV 1259
Db      1194 PEYLTPQGAAPQPHPPAPAFSPAFDNLVYWDQDPPEBGAPEPSTFKCTPTAENPEYLGIDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 12
AAE20479
ID      AAE20479 standard; Protein; 1255 AA.
XX      AAE20479;
AC      AAE20479;
DT      01-JUL-2002 (first entry)
XX      01-JUL-2002 (first entry)
DE      Human Her-2/neu protein.
XX      Human Her-2/neu protein.
KW      Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
XX      human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
OS      Homo sapiens.
XX      Homo sapiens.
FH      Key
FT      Region
PN      Location/Qualifiers
XX      1021..1030
XX      /note="Naturally processed HLA-B44-restricted epitope"
XX      WO200214503-A2.

```

```

PD      21-FEB-2002.
XX      14-AUG-2001; 2001WO-US41733.
PF      14-AUG-2001; 2000US-225152P.
XX      14-AUG-2000; 2000US-225152P.
PR      28-SEP-2000; 2000US-236428P.
PR      21-FEB-2001; 2001US-270520P.
XX      (CORI-) CORIXA CORP.
PA      Hand-Zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI      McNeill PD, Vedral TS;
XX      WPI; 2002-280758/32.
DR      N-PSDB; AAD32743.
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT      prevention and diagnosis of cancer, preferably breast cancer
XX      disclosure; Page 114-117; 129pp; English.
PS      The invention relates to an isolated Her-2/Neu polypeptide composition
XX      effective for eliciting an immune response. The invention is useful for
CC      eliciting an immune response in a patient, where the patient is human
CC      leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC      The composition is useful for the therapy and diagnosis of cancer,
CC      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC      and other compositions for the diagnosis, prevention and treatment of
CC      human malignancies, for stimulating and/or expanding T cells specific for
CC      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC      patient. The invention is useful for stimulating a T cell response in a
CC      human patient, as probe or primer for nucleic acid hybridisation, to
CC      selectively form duplex molecules with complementary stretches of the
CC      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC      length gene from a suitable library, and to direct expression of a
CC      polypeptide in appropriate host cells. The composition is useful in
CC      prophylactic or therapeutic applications and for the treatment of cancer,
CC      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC      associated malignancies. The invention is useful in gene therapy. The
CC      present sequence is human Her-2/neu protein.
XX      Sequence 1255 AA;
SQ
Query Match      96.9%; Score 6641; DB 23; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
Qy      1 MELALCRWGLLALLPPGAASVQCTGTDKRLPASPEHLDMLRHLVGGQVQGNL 60
Db      1 MELALCRWGLLALLPPGAASVQCTGTDKRLPASPEHLDMLRHLVGGQVQGNL 60
Qy      61 ELTYIPTNASISFLDIOEVQGYVLIANOVROVPLQRLIVRGTOLEFEDYALAVLDNG 120
Db      61 ELTYIPTNASISFLDIOEVQGYVLIANOVROVPLQRLIVRGTOLEFEDYALAVLDNG 120
Qy      121 DPLNNTTPVTGASPGGLREIQRLSTELTKSGVLIQRNPQCYQDTIMKQIFKKNOLA 180
Db      121 DPLNNTTPVTGASPGGLREIQRLSTELTKSGVLIQRNPQCYQDTIMKQIFKKNOLA 180
Qy      181 LTLIDTNSRACHPCSPCKSGRCWGBSSSDQSLRTVCAGGACRCGRLPTDCHEQC 240
Db      181 LTLIDTNSRACHPCSPCKSGRCWGBSSSDQSLRTVCAGGACRCGRLPTDCHEQC 240
Qy      241 AAGCGPRHSPCLCLPHNHSIGELCPALVYNTDTFESMPDEGTYTGASCVNACP 300
Db      241 AAGCGPRHSPCLCLPHNHSIGELCPALVYNTDTFESMPDEGTYTGASCVNACP 300
Qy      301 YNYLSTDVGSCTVCPHNDVETADGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db      301 YNYLSTDVGSCTVCPHNDVETADGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Qy      361 ELFPAGCKITFGSLAPFESFDGDPASNTAPLQEQVQVETLEETITGYLYISWPSLP 420
Db      361 ELFPAGCKITFGSLAPFESFDGDPASNTAPLQEQVQVETLEETITGYLYISWPSLP 420

```

Dp	361	QEEFAGGCKIKFGSLAFLEPSPDGDPASTAPLQEPQLQVFTELBETIGVLYIASMPDLP	420
Qy	421	DLSPFQNIQVIRGRIRLHNGAYSLLTQIGIGISMLGRISRELGSGALLHNHTHLCFNTV	480
Dp	421	DLSPFQNIQVIRGRIRLHNGAYSLLTQIGIGISMLGRISRELGSGALLHNHTHLCFNTV	480
Qy	481	PMDQLEFRPHQALLHTTARPEDECVGEGLACHOLCARHCWGCPGTQCVCNCSQFLRGEC	540
Dp	481	PMDQLEFRPHQALLHTTARPEDECVGEGLACHOLCARHCWGCPGTQCVCNCSQFLRGEC	540
Qy	541	VEECGVLQGLFREYVNAHCLPCHPECCQPNQSVTCFGRPADQCYACAHYKDPFCVARC	600
Dp	541	VEECGVLQGLFREYVNAHCLPCHPECCQPNQSVTCFGRPADQCYACAHYKDPFCVARC	600
Qy	601	PSGVKPDLSYPIWKEPDEBEGACQCPINCHSCVDLDDKCAPAQRASPLSTISAVAG	660
Dp	601	PSGVKPDLSYPIWKEPDEBEGACQCPINCHSCVDLDDKCAPAQRASPLSTISAVAG	660
Qy	661	ILLVVLGVVIGILIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRLIKEFNN	720
Dp	661	ILLVVLGVVIGILIKRRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAMRLIKEFNN	719
Qy	721	FTVSWLRAVPKYSAS-HLEITYXKGWIDGGEVWKIPVAIKYLRENTSRKAKELIDEAYV	779
Dp	720	-----LKKVNLVSGAGAGTYKGIWIDGGEVWKIPVAIKYLRENTSRKAKELIDEAYV	773
Qy	780	MAGVSPVYVSRLLIGLCITSTVOLVQLMPYGCILDHVENRGRLSQDOLLNMCQIAGM	839
Dp	774	MAGVSPVYVSRLLIGLCITSTVOLVQLMPYGCILDHVENRGRLSQDOLLNMCQIAGM	833
Qy	840	SYLEDDRVLVRDLAARNVLVKS PNHVKITDPELARLDIDETEHYADGKVPDKMALES	899
Dp	834	SYLEDDRVLVRDLAARNVLVKS PNHVKITDPELARLDIDETEHYADGKVPDKMALES	893
Qy	900	ILRRFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICITDIYM	959
Dp	894	ILRRFRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICITDIYM	953
Qy	960	IMVKCMIIDSECRPFRFELVSEFSMARDPORFVVIQNEIDLGPA SPLDSTFYRSLLEDD	1019
Dp	954	IMVKCMIIDSECRPFRFELVSEFSMARDPORFVVIQNEIDLGPA SPLDSTFYRSLLEDD	1013
Qy	1020	MGLDVADEEYLVPOQGFPCPDPAAGAGMWHHRHSSSTRSGGDLTGLSESEEARPS	1079
Dp	1014	MGLDVADEEYLVPOQGFPCPDPAAGAGMWHHRHSSSTRSGGDLTGLSESEEARPS	1073
Qy	1080	PLASEGAGSDVFDDDLGMGAKGLQSLPTHDPSFLQVSEDPYPLPSETDGYAAPLTC	1133
Dp	1074	PLASEGAGSDVFDDDLGMGAKGLQSLPTHDPSFLQVSEDPYPLPSETDGYAAPLTC	1133
Qy	1140	SPQPEYVNPQDVRPOPSPSBRBGLPLAARPACATLERAKTILSPGKGVVYKDVAFAGCAVEN	1199
Dp	1134	SPQPEYVNPQDVRPOPSPSBRBGLPLAARPACATLERAKTILSPGKGVVYKDVAFAGCAVEN	1193
Qy	1200	PEYLTLPQGAAPQPPAPSPAFPNLTYWDDPPERCAPSPSTFKGTPTAENPEYLVGIDV	1255
Dp	1194	PEYLTLPQGAAPQPPAPSPAFPNLTYWDDPPERCAPSPSTFKGTPTAENPEYLVGIDV	1255
Qy	1260	PV 1261	
Dp	1254	PV 1255	
RESULT 13			
AAMS1143			
ID	AAMS1143	standard; Protein; 1255 AA.	
XX	AAMS1143;		
XX	AC		
XX	DT	17-JUN-2002 (first entry)	
DE	Human Her-2/neu oncogene-encoded p185 glycoprotein.		
XX			

```

QY 61 ELTYLPTNASLSFQDIOEVGVYLIANQVRYVFLQRLRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFQDIOEVGVYLIANQVRYVFLQRLRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASFGSLRELOLRSLTEILKGVLIQNPOLCYODTILMKDIFHKNOLA 180
Db 121 DPLNNTPTVTGASFGSLRELOLRSLTEILKGVLIQNPOLCYODTILMKDIFHKNOLA 180
QY 181 LTLIDTNRSRACHPCSPMCKGSRGWSSESDCQSLTFTVACGACRCKGPLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGWSSESDCQSLTFTVACGACRCKGPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLCLPHNHSICELHCPALVTNTQTPESMPBERGYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLCLPHNHSICELHCPALVTNTQTPESMPBERGYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHNEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360
QY 361 ELFEAGCKKIFGSLAFLEPESFDGDPASTAPLOEQOVPELTLEITGYLISAMPDLSL 420
Db 361 ELFEAGCKKIFGSLAFLEPESFDGDPASTAPLOEQOVPELTLEITGYLISAMPDLSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLRELSGSLALIHNTHLCEVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLRELSGSLALIHNTHLCEVHTV 480
QY 481 PMDOLFRNPHOALHTANRPDECEVGBGLACHOLCAFGHCMPGPTQCVNCSQFLRQEC 540
Db 481 PMDOLFRNPHOALHTANRPDECEVGBGLACHOLCAFGHCMPGPTQCVNCSQFLRQEC 540
QY 541 VEEGRVLOGLPREYVNRHCLPCHPECOPOKGSVTCGPEADQCAAHYKDPFCVARC 600
Db 541 VEEGRVLOGLPREYVNRHCLPCHPECOPOKGSVTCGPEADQCAAHYKDPFCVARC 600
QY 601 PSYKPDLSYMPIMKPFDEBACQPCPINCTHSCVDLDKCPAEQASPLTISVAVVG 660
Db 601 PSYKPDLSYMPIMKPFDEBACQPCPINCTHSCVDLDKCPAEQASPLTISVAVVG 660
QY 661 ILVYVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSAMNQOMRLKEFNN 720
Db 661 ILVYVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSAMNQOMRLKEFNN 720
QY 721 FTVSFMLRVPKVSAS-HLETYYKGIWIPDGENVYKIPAIKYLRENTSPPKANKETLDEAYV 779
Db 721 FTVSFMLRVPKVSAS-HLETYYKGIWIPDGENVYKIPAIKYLRENTSPPKANKETLDEAYV 779
QY 780 MAGVGSPPYVSRLLGICLTSTYQVLTQMLPYGCLLDHVRNRRGRIGSODLLNMCQIAKGM 839
Db 780 MAGVGSPPYVSRLLGICLTSTYQVLTQMLPYGCLLDHVRNRRGRIGSODLLNMCQIAKGM 839
QY 839 774 MAGVGSPPYVSRLLGICLTSTYQVLTQMLPYGCLLDHVRNRRGRIGSODLLNMCQIAKGM 839
Db 839 774 MAGVGSPPYVSRLLGICLTSTYQVLTQMLPYGCLLDHVRNRRGRIGSODLLNMCQIAKGM 839
QY 840 SYLEDVRLVHRDLAARVVLKSPHVKITDFGLARLIDIDETEHADGGVPIKXMALES 899
Db 840 SYLEDVRLVHRDLAARVVLKSPHVKITDFGLARLIDIDETEHADGGVPIKXMALES 899
QY 899 834 SYLEDVRLVHRDLAARVVLKSPHVKITDFGLARLIDIDETEHADGGVPIKXMALES 899
Db 899 834 SYLEDVRLVHRDLAARVVLKSPHVKITDFGLARLIDIDETEHADGGVPIKXMALES 899
QY 900 ILRRRFTHSDVMSYGVYTWELMTFGAKPYDGPAREIPDLLEKGERLPPPICTIDVYM 959
Db 900 ILRRRFTHSDVMSYGVYTWELMTFGAKPYDGPAREIPDLLEKGERLPPPICTIDVYM 959
QY 959 894 ILRRRFTHSDVMSYGVYTWELMTFGAKPYDGPAREIPDLLEKGERLPPPICTIDVYM 959
Db 959 894 ILRRRFTHSDVMSYGVYTWELMTFGAKPYDGPAREIPDLLEKGERLPPPICTIDVYM 959
QY 960 IMVCKMWIDECRRREFELVSEPSRMARDPORFVIONEDLGPASPLDSTFFYSLLDEDD 1019
Db 960 IMVCKMWIDECRRREFELVSEPSRMARDPORFVIONEDLGPASPLDSTFFYSLLDEDD 1019
QY 1019 954 IMVCKMWIDECRRREFELVSEPSRMARDPORFVIONEDLGPASPLDSTFFYSLLDEDD 1019
Db 1019 954 IMVCKMWIDECRRREFELVSEPSRMARDPORFVIONEDLGPASPLDSTFFYSLLDEDD 1019
QY 1020 MGDVLDAEYLVPCQGFCECDPAFAGAGVYHRRHSSSTRSGGDLTLGLEPSEEBEAPRS 1079
Db 1020 MGDVLDAEYLVPCQGFCECDPAFAGAGVYHRRHSSSTRSGGDLTLGLEPSEEBEAPRS 1079
QY 1079 1014 MGDVLDAEYLVPCQGFCECDPAFAGAGVYHRRHSSSTRSGGDLTLGLEPSEEBEAPRS 1079
Db 1079 1014 MGDVLDAEYLVPCQGFCECDPAFAGAGVYHRRHSSSTRSGGDLTLGLEPSEEBEAPRS 1079
QY 1080 PLAPBSGASDVDPGDI-GMGAAKGLQSLPTHDSPLQRYSEDPVPLPSTTDGYVAPLTC 1139
Db 1080 PLAPBSGASDVDPGDI-GMGAAKGLQSLPTHDSPLQRYSEDPVPLPSTTDGYVAPLTC 1139
QY 1139 1074 PLAPBSGASDVDPGDI-GMGAAKGLQSLPTHDSPLQRYSEDPVPLPSTTDGYVAPLTC 1139
Db 1139 1074 PLAPBSGASDVDPGDI-GMGAAKGLQSLPTHDSPLQRYSEDPVPLPSTTDGYVAPLTC 1139
QY 1140 SPOEYVNOPDVAPQPSPREGFLPARPAGATLERAKTUSPGKNGVAVKDVAFGAVEN 1199
Db 1140 SPOEYVNOPDVAPQPSPREGFLPARPAGATLERAKTUSPGKNGVAVKDVAFGAVEN 1199

```

```

Db 1134 SPOEYVNOPDVAPQPSPREGFLPARPAGATLERAKTUSPGKNGVAVKDVAFGAVEN 1199
QY 1200 PEXYLTQGGAAQPPHPPAFSPAFDLVYWDODPPERGAPSTFKGTPTAENPEYGLDV 1259
Db 1194 PEXYLTQGGAAQPPHPPAFSPAFDLVYWDODPPERGAPSTFKGTPTAENPEYGLDV 1253
QY 1260 PV 1261
Db 1254 PV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein, 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytosolic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AMU; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001MO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
XX
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR MPI: 2002-280741/32.
XX
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure: Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC consisting of an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA:
XX
Query Match 96.9%; Score 6641; DB 23; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
QY 1 MELALCRWGLLALLPPGAASVYCTGDMKRLPASPETHLDMRLHYGCGVYQGNL 60
Db 1 MELALCRWGLLALLPPGAASVYCTGDMKRLPASPETHLDMRLHYGCGVYQGNL 60
QY 61 ELTYLPTNASLSFQDIOEVGVYLIANQVRYVFLQRLRVGTQLFEDNYALAVLDNG 120

```

Db	61	ELTYLPTNMSLSFIQDIOEVGVYLIANQVRQVPLQRLIRVTRQQLFEDNYALAVLDNG	120
Qy	121	DPLNNTTPVTGASPGGLRELOJRSUTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA	180
Db	121	DPLNNTTPVTGASPGGLRELOJRSUTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA	180
Qy	181	LTLIDTNSRACHPCSPCKGRCMGESSEDCOSLTRIVCAGGACRCKGPLPTDCCHQC	240
Db	181	LTLIDTNSRACHPCSPCKGRCMGESSEDCOSLTRIVCAGGACRCKGPLPTDCCHQC	240
Qy	241	AAGCTGPKISDCLAFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCTTAC	300
Db	241	AAGCTGPKISDCLAFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCTTAC	300
Qy	301	YNYLSTDVSGCTIVCEPLNCEVTAEDGTORCEKSKPCARVCYGLGSHLEVRVATSAN	360
Db	301	YNYLSTDVSGCTIVCEPLNCEVTAEDGTORCEKSKPCARVCYGLGSHLEVRVATSAN	360
Qy	361	ELFPAQCKKIFGSLAFLEPESFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLP	420
Db	361	IOEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLP	420
Qy	421	DLSVQNLQVIRGRILNHGASLTLOGLISMLGRSLRELGSGLALIHNTHLCFVHTV	480
Db	421	DLSVQNLQVIRGRILNHGASLTLOGLISMLGRSLRELGSGLALIHNTHLCFVHTV	480
Qy	481	PMDOLFENPHQALHTANRPEDECEVGEGLACHOLCARHCWGPPTOCVNSQPLRQEC	540
Db	481	PMDOLFENPHQALHTANRPEDECEVGEGLACHOLCARHCWGPPTOCVNSQPLRQEC	540
Qy	541	VEECKRVLOGLEFREYNARHCLPCHPECOPOKSGVTCGPEADQCVACAHYDPPFCVAC	600
Db	541	VEECKRVLOGLEFREYNARHCLPCHPECOPOKSGVTCGPEADQCVACAHYDPPFCVAC	600
Qy	601	PSGVKPDSTYMPINKFPEEGACQPCINCHSCVDLDDKCPREQASPLTISAVG	660
Db	601	PSGVKPDSTYMPINKFPEEGACQPCINCHSCVDLDDKCPREQASPLTISAVG	660
Qy	661	ILVVVIGVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAMRLKEFN	720
Db	661	ILVVVIGVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAMRLKEFN	720
Qy	721	FTVSFWLAVPRVSA-SILETYVKGIMTPDGENVMIPIVAKLRENTSPKAKETLDAVY	779
Db	720	-----LKKVRLGSGAFVYKGIWIPDGENVMIPIVAKLRENTSPKAKETLDAVY	773
Qy	780	MAGVSPYVSRLLGLCLTSTVQVLTQMLPYGCLLDHYENRGRGSDLLNMCQIAKGM	839
Db	774	MAGVSPYVSRLLGLCLTSTVQVLTQMLPYGCLLDHYENRGRGSDLLNMCQIAKGM	833
Qy	840	SYLEPVRLVHDLAARNLVKSPNHVKITDFGLARLIDIDETVHADGKVPIMMALES	899
Db	834	SYLEPVRLVHDLAARNLVKSPNHVKITDFGLARLIDIDETVHADGKVPIMMALES	893
Qy	900	ILRRFTHQSDVMSYGVTVMEIMTFGAKPYDIPAREIPDLLEKGERLPPOPCITIVYM	959
Db	894	ILRRFTHQSDVMSYGVTVMEIMTFGAKPYDIPAREIPDLLEKGERLPPOPCITIVYM	953
Qy	960	IMVCKWMTIDSECRPFRELVSEFSAMARDPQRFVVIQNEDELGPASPLDSTYRSLDDDD	1019
Db	954	IMVCKWMTIDSECRPFRELVSEFSAMARDPQRFVVIQNEDELGPASPLDSTYRSLDDDD	1013
Qy	1020	MGDLVDAEYLVPOGGFCPPDPAAGAGMTHRRSSSTRSGGDLITLGLPSEBEAPRS	1079
Db	1014	MGDLVDAEYLVPOGGFCPPDPAAGAGMTHRRSSSTRSGGDLITLGLPSEBEAPRS	1073
Qy	1080	PLAPSEGASDVFDGDLGMAAKGISO.LPTHDPSPLOKYSDDFTVPLPSETDGYVAP.LTC	1139
Db	1074	PLAPSEGASDVFDGDLGMAAKGISO.LPTHDPSPLOKYSDDFTVPLPSETDGYVAP.LTC	1133
Qy	1140	SPOPEYVNCQVPRPOPFSPRSGPLPAARFAGATLERAKTSLPGKNGVQVDFAFGAVEN	1199
Db	1134	SPOPEYVNCQVPRPOPFSPRSGPLPAARFAGATLERAKTSLPGKNGVQVDFAFGAVEN	1193
Qy	1200	PEYLTPOGGAAPQPPPPAFPAFNTLVYDODDPEERGAPESTFGPTAENPEYLGUDV	1259
Db	1194	PEYLTPOGGAAPQPPPPAFPAFNTLVYDODDPEERGAPESTFGPTAENPEYLGUDV	1253
Qy	1260	PV 1261	
Db	1254	PV 1255	
RESULT 15			
AA39568			
ID	AA39568 standard; Protein; 1433 AA.		
XX			
AC	AA39568;		
XX			
DT	07-FEB-1994 (first entry)		
XX			
DE	Sequence of c-erbB-2 tumour antigen.		
XX			
KM	Tumour antigen; c-erbB-2; glycoprotein.		
XX			
OS	Homo sapiens.		
XX			
PN	MO9316185-A.		
PD	19-AUG-1993.		
XX			
PF	05-FEB-1993; 93MO-US01055.		
XX			
PR	06-FEB-1992; 92US-0831967.		
XX			
PA	(CETU) CETUS ONCOLOGY CORP.		
XX			
PA	(CREA-) CREATIVE BIOMOLECULES INC.		
XX			
PI	Houston LL, Huston JS, Oppermann H, Ring DB;		
XX			
DR	WPI; 1993-272889/34.		
XX			
DR	N-FSDB; AAQ46083.		
XX			
PT	New single chain Fv polypeptide binding to C-erbB-2 tumour		
XX			
PT	antigen - for imaging or treating breast or ovarian cancer etc.		
XX			
PS	Disclosure; pages 48-54; 87pp; English.		
XX			
CC	c-erbB-2 refers to a protein antigen expressed on the surface of		
XX			
CC	tumour cells, such as breast and ovarian tumour cells, which is an		
XX			
CC	approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric		
XX			
CC	pt. of about 5.3 (see AAQ46083, AA39568). X in AA39563 represents		
XX			
CC	the location of a stop codon in AAQ46083.		
XX			
SQ	Sequence 1433 AA;		
Query Match 96.3%; Score 6598; DB 14; Length 1433;			
Best Local Similarity 96.8%; Pred. No. 0;			
Matches 1221; Conservative 7; Mismatches 26; Indels 8; Gaps 2;			
Qy	1	MELAAICRMGLIALILPPGAASTOVCTGDMKRLRPASPEHILDMRLHYOQVQGNL	60
Db	1	MELAAICRMGLIALILPPGAASTOVCTGDMKRLRPASPEHILDMRLHYOQVQGNL	60
Qy	61	ELTYLPTNMSLSFIQDIOEVGVYLIANQVRQVPLQRLIRVTRQQLFEDNYALAVLDNG	120
Db	61	ELTYLPTNMSLSFIQDIOEVGVYLIANQVRQVPLQRLIRVTRQQLFEDNYALAVLDNG	120
Qy	121	DPLNNTTPVTGASPGGLRELOJRSUTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA	180
Db	121	DPLNNTTPVTGASPGGLRELOJRSUTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA	180
Qy	181	LTLIDTNSRACHPCSPCKGRCMGESSEDCOSLTRIVCAGGACRCKGPLPTDCCHQC	240
Db	181	LTLIDTNSRACHPCSPCKGRCMGESSEDCOSLTRIVCAGGACRCKGPLPTDCCHQC	240

QY 241 AACCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTESMENPEGRYTFGASCTYACP 300
Db 241 AACCTGPKHSDCLACLPNHSIGICELHCPALVTYNTDTESMENPEGRYTFGASCTYACP 300
QY 301 YNYLSTDVSGCTLVCPHNGEVAETADGTORCEKSKPCARVCYGLMOYIKANSKFIGIT 360
Db 301 YNYLSTDVSGCTLVCPHNGEVAETADGTORCEKSKPCARVCYGLMOYIKANSKFIGIT 360
QY 361 ELEFFAGCKKIFGSLAFLPESFDDPASNTPALQPEQOVETLEETGYLYISAMPDLP 420
Db 361 IOEFAGCKRIFGSLAFLPESFDDPASNTPALQPEHLOVETLEETGYLYISAMPDLP 420
QY 421 DLSVFQNLQYIRGRILHNGAYSLTQGLGISMLGLSLRLSGLALIHNNHLCVHTV 480
Db 421 DLSVFQNLQYIRGRILHNGAYSLTQGLGISMLGLSLRLSGLALIHNNHLSFHTV 480
QY 481 PMQOLFPNPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPFOCVNCSOFLRGEC 540
Db 481 PMQOLFRNPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPFOCVNCSOFLRGEC 540
QY 541 VEECRVLOGLPREYVASHCLPCHPECPQNGSVTCFGBADQCAAHYKDPFCVAC 600
Db 541 VEECRVLOGLPREYVASHCLPCHPECPQNGSVTCFGBADQCAAHYKDPFCVAC 600
QY 601 PSQVPPDLSYMPIWKPFDEGACOPINCTHSCVDLDKCPAEORASPLTSIVSAVVG 660
Db 601 PSQVPPDLSYMPIWKPFDEGACOPINCTHSCVDLDKCPAEORASPLTSIVSAVVG 660
QY 661 ILLVVLGVVEGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAQRILKEFNN 720
Db 661 ILLVVLGVVEGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAQRILKETE 719
QY 721 FTVSFMLRVKVSAS-HLETVYKGIWIPGSENVKIPYALIVLRENTSPKANKEILDEAYV 779
Db 721 FTVSFMLRVKVSAS-HLETVYKGIWIPGSENVKIPYALIVLRENTSPKANKEILDEAYV 779
QY 779 720 -----LRKVKVIGSSGFGTVYKGIWIPGSENVKIPYALIVLRENTSPKANKEILDEAYV 773
Db 779 720 -----LRKVKVIGSSGFGTVYKGIWIPGSENVKIPYALIVLRENTSPKANKEILDEAYV 773
QY 780 MAGVSPYVSRLLGICITSTVQVLTQIMPYGCLLDHYRENRGLSGQDLINMCMQIAKGM 839
Db 774 MAGVSPYVSRLLGICITSTVQVLTQIMPYGCLLDHYRENRGLSGQDLINMCMQIAKGM 833
QY 840 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALES 899
Db 834 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALES 893
QY 900 ILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKGERLPQPPICITDVM 959
Db 894 ILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLEKGERLPQPPICITDVM 953
QY 960 IMVKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEBGLPASPLDSTFYASLLEDD 1019
Db 954 IMVKCMWIDSECRPRRELVSSESRMARDPQRFVVIQNEBGLPASPLDSTFYASLLEDD 1013
QY 1020 MGLVDAEELVLPQOGFCPPDPAFGAGMWHHRSSSTRSGGDLITLGLBSEEEAPRS 1079
Db 1014 MGLVDAEELVLPQOGFCPPDPAFGAGMWHHRSSSTRSGGDLITLGLBSEEEAPRS 1073
QY 1080 PLAPSEGAGSDVDGLGMAAKGLOSLPTHDPSPLOQYSEDPVLPSETDGYVAPLTC 1139
Db 1074 PLAPSEGAGSDVDGLGMAAKGLOSLPTHDPSPLOQYSEDPVLPSETDGYVAPLTC 1133
QY 1140 SPQPEYVNDPVPQPPSPREGPLPAARPAATLERAKTILSPKNGVVKVPAFGAVEN 1199
Db 1134 SPQPEYVNDPVPQPPSPREGPLPAARPAATLERAKTILSPKNGVVKVPAFGAVEN 1193
QY 1200 PEYLTPQGAAPQPHPPAPSPAFNDLYYWDQDPPERGAAPPSTFKGTPTAENPEYGLDV 1259
Db 1194 PEYLTPQGAAPQPHPPAPSPAFNDLYYWDQDPPERGAAPPSTFKGTPTAENPEYGLDV 1253
QY 1260 PV 1261
Db 1254 PV 1255

Search completed: July 22, 2003, 09:16:49
Job time : 42.1104 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds

(without alignments)
5413.772 Million cell updates/sec

Title: SEQ4-710-730-14

Sequence: 1 MELALCRWGLLALLPPCA.....TFKGTPTAENPEYIGIDVY 1255

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6602	96.9	1255	1 A24571	protein-tyrosine k
2	5805	85.2	1260	1 TVRNU	protein-tyrosine k
3	5795.5	85.1	1254	2 I48161	p-185 precursor
4	3035	44.5	1210	1 GQHB	epidermal growth f
5	3006	44.1	1210	2 A53183	epidermal growth f
6	2983.5	43.8	1223	1 TVCHLV	epidermal growth f
7	2852.5	41.9	1308	2 A47253	epidermal growth f
8	2559	37.6	1166	1 S06142	protein-tyrosine k
9	2307.5	33.9	1342	2 A36923	kinase-related tra
10	2321.5	32.6	1339	2 TC4827	epidermal growth f
11	1655.5	24.3	688	1 TVPLV	epidermal growth f
12	1592	23.4	604	1 TVYUH	protein-tyrosine k
13	1541	22.6	544	2 S35745	protein-tyrosine k
14	1534	22.5	545	2 S00727	protein-tyrosine k
15	1517	22.3	540	2 B44776	kinase-related tra
16	1515	22.2	540	1 TVPEB	protein-tyrosine k
17	1514.5	22.2	1330	1 GQFPE	epidermal growth f
18	1509	22.1	644	2 A36325	epidermal growth f
19	1217	17.9	1333	2 E88257	protein let-23 (lm
20	1217	17.9	1374	2 S70712	protein-tyrosine k
21	1131	16.6	1369	2 S70713	protein-tyrosine k
22	1126	16.5	527	2 A42032	epidermal growth f
23	1114	16.4	1717	1 A45558	epidermal growth f
24	975.5	14.3	843	2 A27331	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth f
26	754.5	11.1	311	2 S18808	protein-tyrosine k
27	693	10.2	1363	2 T43220	insulin-like growt
28	657	9.6	1382	1 INHUR	insulin receptor p
29	656.5	9.6	1372	2 A34157	insulin receptor p

30	653	9.6	1383	2 A36080	insulin receptor p
31	643.5	9.4	1300	2 A36502	insulin receptor-r
32	632	9.3	1477	2 T18534	protein-tyrosine k
33	630	9.2	1607	2 T43212	insulin-like growt
34	629	9.2	1268	2 B36502	insulin receptor-r
35	602	8.8	1367	1 TGHUR1	insulin-like growt
36	583	8.6	1371	2 A33837	insulin-like growt
37	581.5	8.5	2148	1 A56081	insulin receptor
38	580	8.5	1114	1 S05582	protein-tyrosine k
39	579	8.5	987	2 A54092	protein-tyrosine k
40	575	8.4	2101	2 S57245	insulin receptor (
41	569	8.4	1390	2 T30346	insulin receptor
42	559	8.2	987	2 T48652	mouse developmenta
43	551.5	8.1	952	2 I50612	protein-tyrosine k
44	548.5	8.1	801	4 TVHUR	transforming prote
45	548.5	8.1	984	2 A39753	protein-tyrosine k

ALIGNMENTS

RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal grow

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epi

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chr

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

A/Residues: 1-517; RALL, 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU>

R/King, C.R.; Kraus, M.H.; Aaronson, S.A. Science 223, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; MUID:85272597; PMID:2992089

A/Accession: I59509

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 832-909 <REX>

A/Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R/Rat, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D. Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio

A/Reference number: I57622; MUID:87268698; PMID:3039351

A/Accession: I57622

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AAA58637.1; PID:9553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetics:
 A:Gene: GDB:ERBB2; NCL; NEU; HER-2
 A:Cross-references: GDB:120613; OMTM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the list of introns is incomplete
 C:Function:
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphinase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:12-125/Domain: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:12-653/Domain: extracellular #status predicted <EXT>
 F:70-304/Domain: EGF receptor extracellular domain repeat <EP1>
 F:395-605/Domain: EGF receptor extracellular domain repeat <EP2>
 F:654-675/Domain: transmembrane #status predicted <INT>
 F:676-1255/Domain: intracellular #status predicted <INT>
 F:726-734/Region: protein kinase homology <KIN>
 F:658-124,187,229,530,571,629/Binding site: carbohydrate (asn) (covalent) #status predicted
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)
 Query Match 96.9%; Score 6602; DB 1; Length 1255;
 Best Local Similarity 96.9%; Pred. No. 4,66-277;
 Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

Dy 1 MELALCGWGLLLLPAGASTQVCTGDKRLRLSPETHLDMRLHYOGGCVGNL 60
 Db 1 MELALCGWGLLLLPAGASTQVCTGDKRLRLSPETHLDMRLHYOGGCVGNL 60
 Qy 61 ELTYLPTNASLSPQIDQEVGVYLAHQVQVPLRLRIVRGTOLEFEDNALAVDNG 120
 Db 61 ELTYLPTNASLSPQIDQEVGVYLAHQVQVPLRLRIVRGTOLEFEDNALAVDNG 120
 Qy 121 DPLNNTTPVYTCASGGLRELELRLTELKGVLIQGNPOLCYDITLMDIPIHNNOLA 180
 Db 121 DPLNNTTPVYTCASGGLRELELRLTELKGVLIQGNPOLCYDITLMDIPIHNNOLA 180
 Qy 121 DPLNNTTPVYTCASGGLRELELRLTELKGVLIQGNPOLCYDITLMDIPIHNNOLA 180
 Db 121 DPLNNTTPVYTCASGGLRELELRLTELKGVLIQGNPOLCYDITLMDIPIHNNOLA 180
 Qy 181 LTLIDNRSRACHPCSPCKGSRGCGESSEDQSLTRVCGGACGKGPLPDDCCEQC 240
 Db 181 LTLIDNRSRACHPCSPCKGSRGCGESSEDQSLTRVCGGACGKGPLPDDCCEQC 240
 Qy 241 AAGCTGPHSCLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
 Db 241 AAGCTGPHSCLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
 Qy 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCEKSKCARCYVGLGMQYIKANSKIFGIT 360
 Db 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCEKSKCARCYVGLGMQYIKANSKIFGIT 360
 Qy 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCEKSKCARCYVGLGMQYIKANSKIFGIT 360
 Db 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCEKSKCARCYVGLGMQYIKANSKIFGIT 360
 Qy 361 ELLEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQVETLEITGYLYISAMPDLP 420
 Db 361 ELLEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQVETLEITGYLYISAMPDLP 420
 Qy 361 IOEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQVETLEITGYLYISAMPDLP 420
 Db 361 IOEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQVETLEITGYLYISAMPDLP 420
 Qy 421 DLSTFQNLQVIRGILHNGAYSLTLOGISWGLRLRLRELGSLAIHNNTHLCFHTYV 480
 Db 421 DLSTFQNLQVIRGILHNGAYSLTLOGISWGLRLRLRELGSLAIHNNTHLCFHTYV 480
 Qy 481 PMDOLFRNFQALHTANRPEDECVEGLACHQACAGHCGWGPPTQCVNCSQFLRQEC 540
 Db 481 PMDOLFRNFQALHTANRPEDECVEGLACHQACAGHCGWGPPTQCVNCSQFLRQEC 540
 Qy 541 VEEGRVYQGLPREVYNAHCLCPHEPCQPOWGSITCGPRADQVAAAHYKDPFCYAR 600
 Db 541 VEEGRVYQGLPREVYNAHCLCPHEPCQPOWGSITCGPRADQVAAAHYKDPFCYAR 600
 Qy 601 PSQVKKPDLSTYMPIMKFPDEGACQPCPCINCHSCVDDDDGCPAEQASPLTSTVSAVVG 660
 Db 601 PSQVKKPDLSTYMPIMKFPDEGACQPCPCINCHSCVDDDDGCPAEQASPLTSTVSAVVG 660

Dy 601 PSQVKKPDLSTYMPIMKFPDEGACQPCPCINCHSCVDDDDGCPAEQASPLTSTVSAVVG 660
 Qy 661 ILLVYLVGVVFGILLKRRQOKIKYTMRLLOETLVEPLPSGAMPNOAQRIILKETEL 720
 Db 661 ILLVYLVGVVFGILLKRRQOKIKYTMRLLOETLVEPLPSGAMPNOAQRIILKETEL 720
 Qy 721 RKVKVLGSAFGFNNFTVFWLRVP----KVSASHLEVRNRENTSPKANKEILDEAYYMA 775
 Db 721 RKVKVLGSAFGF---TVYKGIWPDGEVHKIPVA-IKVLRNRENTSPKANKEILDEAYYMA 775
 Qy 776 GVGSPVSVSLGILCTSTVQVLTQMLPGVCLLDHRENRGRSGDILLNMCQIAKMSY 835
 Db 776 GVGSPVSVSLGILCTSTVQVLTQMLPGVCLLDHRENRGRSGDILLNMCQIAKMSY 835
 Qy 836 LEDVRVYHDDLAARVYLVSPHNVKITDGLARLLDIBETHEAAGKPIKMALESL 895
 Db 836 LEDVRVYHDDLAARVYLVSPHNVKITDGLARLLDIBETHEAAGKPIKMALESL 895
 Qy 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPQPICTIDVYIM 955
 Db 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIREIPDLLEKGERLPQPICTIDVYIM 955
 Qy 956 VKCWMIDSECRPRFELVSESRMARDPQRFVIONEDIGPASPLDSTFYSLDEDDMG 1015
 Db 956 VKCWMIDSECRPRFELVSESRMARDPQRFVIONEDIGPASPLDSTFYSLDEDDMG 1015
 Qy 1016 DIVDAEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEPSEEAAPRSL 1075
 Db 1016 DIVDAEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEPSEEAAPRSL 1075
 Qy 1076 ABSEAGSVDFPDGDMGAKLOSLPTIDSEPLQRYSDPVPVLPSETDGYAAPTCTSP 1135
 Db 1076 ABSEAGSVDFPDGDMGAKLOSLPTIDSEPLQRYSDPVPVLPSETDGYAAPTCTSP 1135
 Qy 1136 QPEYVQNPVPRQPSREPRGPPAPARAGATLERAKTISPQNGVYKDFAGAVENE 1195
 Db 1136 QPEYVQNPVPRQPSREPRGPPAPARAGATLERAKTISPQNGVYKDFAGAVENE 1195
 Qy 1196 YLTPQGAAPQHPPPAPFANFNYMDQDPPERGAPPTFKPTPAENPEYGLDVPV 1255
 Db 1196 YLTPQGAAPQHPPPAPFANFNYMDQDPPERGAPPTFKPTPAENPEYGLDVPV 1255

RESULT 2
 TETRNU
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Barmann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein
 A:Reference number: A24562; MUID:86118662; PMID:945311
 A:Accession: A24562
 A:Molecule type: mRNA
 A:Residues: 1-1260 <BAR>
 A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:955746
 R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohn
 Carcinogenesis 12, 1975-1978, 1991
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals:
 2-thiazolyl]formamide or N-methyl-N-nitrosourea.
 A:Reference number: A61204; MUID:92035293; PMID:1682063
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663, 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 C:Genetics:
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

8-680/Domain: transmembrane #status predicted <TMN>
 3-588/Domain: protein kinase homology <KIN>
 1-739/Region: protein kinase ATP-binding motif
 1-191,263,535,576,634/Binding site: carbohydrate (Asn) #status predicted
 11/Binding site: phosphate (Thr) (covalent) #status predicted
 38/Active site: lys #status predicted
 32,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.2%; Score 5805; DB 1; Length 1260;
 est Local Similarity 85.3%; Pred. No. 9,4e-243; Indels 12; Gaps 5;
 atches 1076; Conservative 58; Mismatches 116;

1 MELALCRWGLLALLPRGASTOVCTGDMKRLPASPEHLDMLRHLVGGCVVQGNL 60
 4 MELAAWCRWGLLALLPRGASTOVCTGDMKRLPASPEHLDMLRHLVGGCVVQGNL 63
 61 ELLYPTNASTSFLODIOEVQGVLLAHNQVQVPLQRLIRYRGQVLEEDNALAVLNG 120
 64 ELLYPTNASTSFLODIOEVQGVLLAHNQVQVPLQRLIRYRGQVLEEDNALAVLNG 123
 121 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI 179
 124 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI 183
 180 ALTLIDNRSRACHPCSPMCKSGCWGSESSDQSLRTVCAGGACRCKGPIPTDCCHQ 239
 184 APVIDITNRSRACHPCSPMCKSGCWGSESSDQSLRTVCAGGACRCKGPIPTDCCHQ 243
 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESSMPNPEGRYTFGASCVTAC 239
 244 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESSMPNPEGRYTFGASCVTAC 303
 300 PYNLSTVDSCTVCLPLAHNOEVLAEEDTORCEKSKFCARVYGLQMGVYKANSKFI 359
 304 PYNLSTVDSCTVCLPLAHNOEVLAEEDTORCEKSKFCARVYGLQMGVYKANSKFI 363
 360 TELEFAGCKKIFGSAFLPESFDGDPASNTAPLOEQLQVFTLEIIGVLYISAMPDL 419
 364 NVQEPFGCKKIFGSAFLPESFDGDPASNTAPLOEQLQVFTLEIIGVLYISAMPDL 423
 420 PDLSTVQNLQVIRGRILHNGANSVLTQIGISMLGRLSRELGSGLALHNNHLCFVHT 479
 424 PDLSTVQNLQVIRGRILHNGANSVLTQIGISMLGRLSRELGSGLALHNNHLCFVHT 483
 480 VPMDOLEFRNPHQALHNSGRPEEDLCVSSGLVCNSLCAHGHCMGPGFTQVCNCSHFLRG 543
 484 VPMDOLEFRNPHQALHNSGRPEEDLCVSSGLVCNSLCAHGHCMGPGFTQVCNCSHFLRG 548
 539 ECVECEGVLOGLPREYVNAHCLPCHPECCQPNQNSVTCFGRPADQCAAHYKDPFCVA 598
 544 ECVECEGVLOGLPREYVNAHCLPCHPECCQPNQNSVTCFGRPADQCAAHYKDPFCVA 603
 599 RCPGSGVPRDISYMPIMKRPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTISYSAV 658
 604 RCPGSGVPRDISYMPIMKRPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTISYSAV 663
 659 VGILLVVLGVFGILLIKRQCKIRKYMRLLOETLEVEPLTSGAMPNAPQAKILKET 718
 664 VGILLVVLGVFGILLIKRQCKIRKYMRLLOETLEVEPLTSGAMPNAPQAKILKET 723
 719 ELRKVVLGSGAFGFGNNFTVSWLAVP-----KVASHLEVLRENTSPKANKEILDEAYV 773
 724 ELRKVVLGSGAFGFGNNFTVSWLAVP-----KVASHLEVLRENTSPKANKEILDEAYV 778
 774 MAGVSGPYVSRLLIGICTSTVQVLTQIMBYGCLLDHVENRGRSLGODLLMCMOIAKAM 833
 779 MAGVSGPYVSRLLIGICTSTVQVLTQIMBYGCLLDHVENRGRSLGODLLMCMOIAKAM 838
 834 SYLEADVLRDLAARNVLYKSPNHKITYDFGLARLIDETETHADGCVPIKMALES 893
 839 SYLEADVLRDLAARNVLYKSPNHKITYDFGLARLIDETETHADGCVPIKMALES 898
 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPORPCTIDVYM 953

DB 899 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPORPCTIDVYM 958
 954 INWKCMIMDSECRPFRELVSFSSMARDPQRFVYIQNEDLGASPLDSTYRSLLDD 1013
 959 INWKCMIMDSECRPFRELVSFSSMARDPQRFVYIQNEDLGASPLDSTYRSLLDD 1018
 1014 MGDVDAEYLVQCGFFCPDPAPGAGVHNRHSSSTRSGGDLTLLESEEBAPPS 1073
 1019 MGDVDAEYLVQCGFFCPDPAPGAGVHNRHSSSTRSGGDLTLLESEEBAPPS 1078
 1074 PLAPSGAGSDVDGDLGMGAAGLQSLPHTDPSPIQRSEDPVLPSETDGVAPLTC 1133
 1079 PLAPSGAGSDVDGDLGMGAAGLQSLPHTDPSPIQRSEDPVLPSETDGVAPLTC 1138
 1134 SPQPEYVQDPVRPQPSPREGLPAPAPGATLERAKTLPKGNVYKDVAFAGAVEN 1193
 1139 SPQPEYVQDPVRPQPSPREGLPAPAPGATLERAKTLPKGNVYKDVAFAGAVEN 1198
 1194 PEYLTPOGGAPOHPAPAFADNLYWDQDPERGAPSTKQPTLENPEYLGADV 1253
 1199 PEYLTPOGGAPOHPAPAFADNLYWDQDPERGAPSTKQPTLENPEYLGADV 1258
 1254 PY 1255
 1259 PY 1260

RESULT 3
 148161
 P-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #ext_change 18-Jun-1999
 C:Accession: 148161
 R:Nakamura, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Gene 140, 251-255, 1994
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Accession: 148161; MUID:94193007; PMID:7998275
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:References: 1-1254 <RES>
 A:Cross-references: GB:D16295; NID:9493236; PID:BA03801.1; PID:9747595
 C:Genetics:
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP
 F:118-983/Domain: protein kinase homology <KIN>
 F:126-734/Region: protein kinase ATP-binding motif

Query Match 85.1%; Score 5795.5; DB 2; Length 1254;
 est Local Similarity 84.9%; Pred. No. 2,4e-242; Indels 11; Gaps 4;
 Matches 1070; Conservative 67; Mismatches 112;

1 MELALCRWGLLALLPRGASTOVCTGDMKRLPASPEHLDMLRHLVGGCVVQGNL 60
 1 MELAAWCRWGLLALLPRGASTOVCTGDMKRLPASPEHLDMLRHLVGGCVVQGNL 63
 61 ELLYPTNASTSFLODIOEVQGVLLAHNQVQVPLQRLIRYRGQVLEEDNALAVLNG 120
 61 ELLYPTNASTSFLODIOEVQGVLLAHNQVQVPLQRLIRYRGQVLEEDNALAVLNG 120
 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI 180
 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI 183
 181 LTLIDNRSRACHPCSPMCKSGCWGSESSDQSLRTVCAGGACRCKGPIPTDCCHQ 240
 181 LTLIDNRSRACHPCSPMCKSGCWGSESSDQSLRTVCAGGACRCKGPIPTDCCHQ 240
 241 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESSMPNPEGRYTFGASCVTAC 300
 241 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTESSMPNPEGRYTFGASCVTAC 303

```

301 IYVISTDVGSTLVCPLIHNQEVTAEDTQCEKCSVP CARVYGLIOMQYIKANSFFIGIT 360
301 IYVISTEYVGSCTTLYVCPINNOEVTAEEDTOCEKCSKARVCYGLGMEHLRGARAITSAN 360
361 ELEPAGCKKIFGSLAF.PSEFDDGPASNTPLPELOLVFETLEETITGLYISAMPDLP 420
361 IOEPAGCKKIFGSLAF.PSEFDGNPSSGIALPLPELOLVFETLEETITGLYISAMPDLSL 420
421 DLSFQNLQVIRGILINCAVSLTLQGLISWIGJARS.RELGSGTAL.IHNTHLCPYHTV 480
421 DLSFQNLQVIRGILINCAVSLTLQGLISWIGJARS.RELGSGTAL.IHNTHLCPYHTV 480
481 PMDLEFNPHQALIHNTANPREDECVEGLAHCQLCARGHCMGPGPTQCVNCSQELRGQEC 540
481 PMDLEFNPHQALIHNTANPREDECVEGLAHCQLCARGHCMGPGPTQCVNCSQELRGQEC 540
541 VESQVTLQGLPREYVNAARCLPCHPECPQONSVTCPFEPAQOCVACAHYMDPPCYARC 600
541 VKESCRWKGJ.PREYVNAARCLPCHPECPQONSVTCPFEPAQOCVACAHYMDPPCYARC 600
601 PSQVXKDLAMPKRWKPEDEGACQPCPINCTHSCVDLDDKQCPAQRASPLTSIVSAVVG 660
601 PSQVXKDLAMPKRWKPEDEGACQPCPINCTHSCVDLDDKQCPAQRASPLTSIVSAVVG 660
661 ILVVVLGVVFGILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAKMILKTEL 720
661 ILFVLVGVVFGILIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAKMILKTEL 720
721 RKRYKVLGSAFCFNNFTVSPMLRVP.---KVASHLVEVRENTSPKXKEIIDEAYVWA 775
721 RKRYKVLGSAFCFNNFTVSPMLRVP.---KVASHLVEVRENTSPKXKEIIDEAYVWA 775
776 GVGSFYVSRLLIGICTSTVQCLVLTQMLRYGCLLDHVRNCRRLCSODLLMWQIAKMSY 835
776 GVGSFYVSRLLIGICTSTVQCLVLTQMLRYGCLLDHVRNCRRLCSODLLMWQIAKMSY 835
836 LEDVRLVHRDLAARNVLVKSFNHYKIDFGLARLLDIDREYVADGKPIKMALESL 895
836 LEDVRLVHRDLAARNVLVKSFNHYKIDFGLARLLDIDREYVADGKPIKMALESL 895
896 RRRFHOSSDWSYGVTWELMTFGAKYDGI.PAREI.PDLLEKGERLP.POPICITDIVYIM 955
896 RRRFHOSSDWSYGVTWELMTFGAKYDGI.PAREI.PDLLEKGERLP.POPICITDIVYIM 955
956 VKCMWIDSEGRPREFRELVESEFSRMARDPQRFVITQWEDIGP.SPLDSITFYSSLLEDDMG 1015
956 VKCMWIDSEGRPREFRELVESEFSRMARDPQRFVITQWEDIGP.SPLDSITFYSSLLEDDMG 1015
1016 DLVDAEYLYVPOQGFCCPPAPAGAGMHHRRSSSTRSGGSDTLIGT.EPEEERAPSP 1075
1016 DLVDAEYLYVPOQGFCCPPAPAGAGMHHRRSSSTRSGGSDTLIGT.EPEEERAPSP 1075
1076 APSEGAGSDVFDDDLDMGAAGI.QSLPTHDSPLPQYSEDPVLP.PSEFTDGYAFITCSP 1135
1076 APSEGAGSDVFDDDLDMGAAGI.QSLPTHDSPLPQYSEDPVLP.PSEFTDGYAFITCSP 1135
1136 QEPYVNAQVPVRPQPS.PRSGPL.PAARPGATLERKKTLSPEKQNVADVAFGCAVBNPE 1195
1136 QEPYVNAQVPVRPQPS.PRSGPL.PAARPGATLERKKTLSPEKQNVADVAFGCAVBNPE 1195
1196 YLVPRGSGASQGH -PPALCPAFNDLXYMDQPSERGSPPMTFESTPTAHNEPVLIGLIDVY 1254
1196 YLVPRGSGASQGH -PPALCPAFNDLXYMDQPSERGSPPMTFESTPTAHNEPVLIGLIDVY 1254

```

R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Ye
kg, P.H. Nature 309, 418-425, 1984

A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
A:Accession: A00641; MUID:84219729; PMID:6328312

A:Molecule type: mRNA

A:Residues: 1-1210 <ULN>

A:Cross-references: EMBL:X00588; NID:g11113; PIDN:CAA25240.1; PID:g757924

A>Note: the authors translated the codon AAG for residue 540 as Asn

P: Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Melino, G.T.; Pastan, I.

Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985

A:Title: Characterization and sequence of the promoter region of the human epidermal grow
A:Reference number: A25772; MUID:85270438; PMID:2991899

A:Accession: A25772

A>Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-29 <ISH>

A:Cross-references: GB:M11234; NID:g181991; PIDN:AAA5370.1; PID:g553272

R: Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.

Oncogene Res. 1, 375-396, 1987

A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification of
A:Reference number: S30024; MUID:88217333; PMID:3329716

A:Accession: S30024

A:Molecule type: DNA

A:Residues: 1-29 <HA2>

A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119

J: Haley, J.D.; Waterfield, M.D.

J. Biol. Chem. 266, 1746-1753, 1991

A:Title: Contributory effects of de Novo transcription and premature transcript terminati
A:Reference number: A38672; MUID:91107677; PMID:1988448

A:Accession: A38672

A:Molecule type: DNA

A:Residues: 1-29 <HML>

A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271

A:Experimental source: Carcinoma cell line A431-7

R: Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Metl
Nature 309, 806-810, 1984

A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs c
A:Reference number: A00642; MUID:84245835; PMID:6330563

A:Accession: A00642

A:Molecule type: mRNA

A:Residues: 728-799; TD, 802-811; R, 813-942 <XUY>

A:Experimental source: epidermoid carcinoma cell line A431

R: Rijn, C.R.; Chen, W.S.; Krulger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.M.

Science 224, 843-848, 1984

A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A:Reference number: A43615; MUID:84196372; PMID:6326261

A:Accession: A43615

A:Molecule type: mRNA

A:Residues: 713-964 <LIN>

A:Experimental source: epidermal carcinoma cell line A431

R: Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.

Biochem. Biophys. Res. Commun. 124, 125-132, 1984

A:Reference number: A23062; MUID:85046483; PMID:6093780

A:Accession: A23062

A:Molecule type: mRNA

A:Residues: 1028-1210 <SIM>

R: Weber, W.; Gili, G.N.; Speise, J.

Science 224, 294-297, 1984

A:Reference number: A05281; MUID:84172183; PMID:6324343

A:Accession: A05281

A:Molecule type: protein

A:Residues: 25-30; S, 32-51; 454-467 <WEB>

R: Russo, M.M.; Lukas, T.J.; Cohen, S.; Staros, J.V.

J. Biol. Chem. 260, 5205-5208, 1985

A:Title: Identification of residues in the nucleotide binding site of the epidermal grow
A:Reference number: A60143; MUID:85182650; PMID:2985580

A:Accession: A60143

A:Molecule type: protein

A:Residues: 740-744; X, 746-747 <RUS>

R: Mroczkowski, B.; Mosig, G.; Cohen, S.

J. Mroczkowski, B.; Mosig, G.; Cohen, S.

ature 309, 270-273, 1994
 Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide
 Reference number: A38023; MUID:8419154; PMID:6352948
 Contents: annotation; receptor activity
 Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 Cell 59, 33-43, 1989
 Title: Functional independence of the epidermal growth factor receptor from a domain
 Reference number: A3331; MUID:9003233; PMID:2790960
 Contents: annotation; internalization; signal
 Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor
 C:Genetics:
 A:Gene: GDB:EGFR
 A:Cross-references: GDB:120610; OMTN:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F1-24/Domain: signal sequence #status predicted <SIG>
 F1-24/Product: EGF receptor #status predicted <EXT>
 F1-25-1210/Domain: extracellular domain repeat <EEL>
 F1-25-645/Domain: EGF receptor extracellular domain repeat <EE2>
 F1-25-300/Domain: EGF receptor extracellular domain repeat <EEL>
 F1-25-668/Domain: transmembrane #status predicted <TM>
 F1-25-1210/Domain: intracellular #status predicted <INT>
 F1-10-975/Domain: protein kinase homology <KIN>
 F1-18-726/Region: protein kinase ATP-binding motif
 F1-99-1046/Region: coated-pit mediated internalization signal
 F1-1047-1210/Region: inhibitory
 F1-128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Aan) (covalent) #status predic
 F1-745/Active site: Lys #status experimental

Query Match 48.5%; Score 3035; DB 1; Length 1210;
 Best Local Similarity 48.7%; Pred. No. 1, se-123;
 Matches 619; Conservative 176; Mismatches 359; Indels 116; Gaps 25;

11 LLLALLPPGAA-STGYCTGTMKRLPASPEHMLDLRLXGCGVQVQNGLELYLPTN 68
 14 LLLALCPASRLSEKKVCGGTSNLTQLGTFEDHFLSLQFMFENVEVLGLLEITVQRN 73
 69 ASLFLQDIQVQGVVILAHQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 128
 74 YDLFLKTIQVQGVVILAHQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 126
 129 VTGASPGRLQELRLSLTEILKGVLTORNPLCYQDTIMKDI FHKQNLALTLIDTR 188
 127 ---ANKGLKELPRNMLDEILHGAVFENPNALCVESIQWRDIVSSDFLSNMSMPQNH 183
 189 SRACHPCSPMKGSRGSESSDQSLRTVCAGGCA-RCKGPIPTDCHEQCAAGCTGP 247
 184 LGSQCKDBSPSCNMGSCWGEENCCQKLTITICAGQCSGRCKSPSDCHNCAAGCTGP 243
 248 KHSQCLAGLHFNHSGICELCPALVTYNTDFESMPNDEGRYTFGACVTAACPYNVSTD 307
 244 RSDSLVCRKFEDEATCKTCTCPMLNPTTYQMVNPEGKYSFATCVKCCPRNVVTD 303
 308 VESCTLVCPHNOVTAEDGTORCEKSCFPCARVCYGLQVQITKNSKFEITTELE-PAG 366
 304 HSCVACGADSYEM-EEDGVKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCKCK 361
 367 CKKIFSLAFSPSPGDPASNTAPLOPEOLOFTELEITGYLISAMPDLSPLVSFQ 426
 362 CTISGDLHLPLPAFFGDSFTPLPDLPOBLDLKVKITIGLLILQAPENRTDLHAF 421
 427 NLQVIRGLHNGAVSLTLOGLISWLSRLSELOSGALLHNHNLCFVHTVPMDQF 486
 422 NLEIRGRTHQHOGLAVVSLNITSIGLSKEISDGVVLSGNKLLCYANTINKKLF 481
 487 RNPQALHTANPEDECVGEGLAGHOLCARHGWGMPGPGQVNCSCQFLRGQCEVCRV 546
 482 GTSGQKTKIISNRNENCKATGVCHALCSPEGCGWPEPFDVSCNNSVGRCDVKCL 541
 547 LOGPREYVNAHCLPCHPECPONGSVTCFGEADQVCAHYKDPFPCVARGSVK 606

542 LEGBREFYENSECIQCHPECLPQAMNITCTGRGENDTQCANYIDGPHCVKTCFAGVWG 601
 607 DLSTYPMKFPPEBAGCQPCPINCSTHSCVYDLDKSCPAEORASPLTSVSAVVG---ILL 663
 602 ENNTL-VWKYADAGHYCHLCHNCTYGCPCGPLESCFTNGKPIF-SIATGVGALLLL 658
 664 VVVLGVVFGILLKRCQKIRKYMRLLOETELVEPLTPSGAMPQAKMILKETELRY 723
 659 VVALGIG---LFRRRHIVKRTLRLLQERELVEPLTPSGAPQALLRLKETEFKKI 715
 724 KVLGSAFGFNNTVFSFWLVPVVSASHLEV---LRENTSPKANEILDEAVVAVGVS 779
 716 KVLGSAFG---TVYKGLMPEGEVKVLPVAIKEIREATSPKANEILDEAVVAVSVDN 771
 780 PYVSLIGLCTSVQVLTQMPYGLLDHRENRRGLSGQDLLNMCQIAKMSYEDV 839
 772 PHVRLIGLCTSVQVLTQMPYGLLDHRENRRGLSGQDLLNMCQIAKMSYEDV 831
 840 RLVHRDLAARVLYKSPNHVKITDFGLALLLDITETVHADGGVPIKMALESILRRF 899
 832 RLVHRDLAARVLYKTPQHVKITDFGLALLLDITETVHADGGVPIKMALESILRRY 891
 900 THQSDVMSYGVTVVEMLTFGAKPYDGI PASEIPPLEKGERLPQPICTIDVYMIWKW 959
 892 THQSDVMSYGVTVVEMLTFGSKPYDGI PASEISSILEKGERLPQPICTIDVYMIWKW 951
 960 MIDSCRRPREFLSEFSRMADEPQFVYIQ-NEDLGASPLDSTFTFSLLEDMDGLY 1018
 952 MIDASRRKREFELIEFSRMADEPQFVYIQ-NEDLGASPLDSTFTFSLLEDMDGLY 1011
 1019 DAEELVPOQGFCCPDPAPAGKGMVHNRSSSTRSGGDLTLGEPSEEARSPPLAS 1078
 1012 DAEELVPOQGFCCPDPAPAGKGMVHNRSSSTRSGGDLTLGEPSEEARSPPLAS 1037
 1079 BEAGSDVFPDGDGMGAAGLQSLTPHDSPLQRYSEDETVLPSET-DGYVAPLTCSEQ 1136
 1038 LSATSN-NSVACTIDRNLQSCPIKEDSFQRYSDPFGALTEDSIDTFL-----PV 1089
 1137 PEYVQPDVQPPSPBREGPLPARPAGATLERAKTLSPGKNGVQVDAFAGAVENPEY 1196
 1090 PEYVQPDVQPPSPBREGPLPARPAGATLERAKTLSPGKNGVQVDAFAGAVENPEY 1138
 1197 L-TPGGAAAPQHPAPAFAPADNLYWYQ-----DP-----PERGAPSPSTFGK 1239
 1139 LNTVQ-----PFCVNSTSDSPAHMQKSHQSLDNPDYQDQFFKKAQKPIFGK 1189
 1240 TPTAENPEYL 1249
 1190 S-TAENAEYL 1198

RESULT 5
 A53183
 epidermal growth factor receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C:Accession: A53183; A43818; S24942; A28941; S45325; I49643
 R:Unetake, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.
 Genes Dev. 8, 399-413, 1994
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF recepto
 A:Reference number: A53183; MUID:94170986; PMID:815255
 A:Accession: A53183
 A:Molecule type: mRNA
 A:Residues: 1-1210 <LUE>
 A:Cross-references: GB:U03425
 R:Aviv, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morée, B.
 Oncogene 6, 673-676, 1991
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding
 A:Reference number: A43818; MUID:91232866; PMID:2030916
 A:Molecule type: mRNA
 A:Residues: 1-714 <AVI>
 A:Cross-references: GB:X59698

R.Eisinger, D.P.; Serreio, G.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heiserman, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941, MUID:88330814, PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009, R:Hisby, M.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X76987, NID:948830, PIDN:CAA55587.1, PID:948831
R:Paria, B.C.; Day, S.K.; Andrews, G.K.; Day, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A:Reference number: 149643, MUID:93126380, PMID:7678348
A:Accession: 149643
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:L06864, NID:9193001, PIDN:AAA53029.1, PID:9567201
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F:1-4/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:697/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.1%, Score 3006, DB 2, Length 1210;
Best Local Similarity 48.6%, Pred. No. 2, 7e-122;
Matches 621, Conservative 169, Mismatches 367, Indels 120, Gaps 27;

QY 11 LLLALLPPGAA--STOVCTGDMKLRPASPETHLDMLRHLYGCGVYVQNLLEITVLPN 68
DB 14 LLLALLCAAGALAEKKKCCGTSNRLLTGTFEDHFLSLQRYKYNCEVLLQGLLEITVQRN 73
QY 69 ASLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTLQEDFNALAVLQNGDPLNNTTP 128
DB 74 YDLFFLKTIOEVAGYVLIANTVERIPLENLQIRGNALVENTYVALILSN----- 124
QY 129 VTGASPGGLRELLRLSLLEILKGVLLQGNPQLCYOTILMKD-----FKNNQALATLI 184
DB 125 -YGNRTGLRELPRNLOELLIGAVRSPNNPILCNMTIQMRDIDVQNVFNMSMD--- 180
QY 185 DTNRSRACHPCSPCKSRGSGESSDDCSLRTVCAGGCA-RCKGPLPTDCCHEGCAAG 243
DB 181 -QSHRSSCPKCDPSCPGSGCGGGEENQKTLKILCAQCSHRGRGSPSDCCHNQCAAG 249
QY 244 CTGPKASDCLAFLPHNSGICELHCPALVTYNTTFESMRPBRITFGASCTYACPNY 303
DB 240 CTGPRESDCLVQCFQDEATCKDTCPPMLYNPTTYMDVNPBEKXVFGATCYKCKPRNY 299
QY 304 LSTPVSCTIVACPLAHNOEVTAEQTCRCEKSKPCAVCYGLGQVYIKASKFIGITELE 363
DB 300 VVTHGSCVACRACGPDYEV--EEDDIRCKCKCDGCRVANGIGIGERK-DTSLINATNIK 357

QY 364 -FAGCKKIFGSLAFLPESPFGDPASNTAPLOEOLOVFETLEITCYLYISAMPDLPDL 422
DB 358 HFKYCTAISGDHLIPVAFKSDSFTRTPDLRELEILKTVKEITFOILLIQAMPDWTDL 417
QY 423 SVFQNLQVIRGLIHNGAVSLTQGLISWGLSLRLSGSLALIHNTHLCPHTVPM 482
DB 418 HAFENLEIRGRTHQGFSLAVVGLINTSLGRLSLKEISDDVITISGRNLCYANTIM 477
QY 483 DQFRPHQALHTANPREDECVGSELACHQACAGHCMPPTQCVNCSQFLRQECVE 542
DB 478 KKLFTGNQKTKIMNRAKDCVAVNHVNCPLSSBGCWGPBRDVSQNVSRGECVE 537
QY 543 ECRVLQGLREYVNAHCLPCHPECOPOGVSYTCFPEADQCVACAHYVDPFCVYRCS 602
DB 538 KNILLEGDEPREFVENSECIGCHPECLPQAMNITTCGRGDNICQCHAYIDGHCVATCA 597
QY 603 GYKPDLSYPIWPKFPBEGACQCPPLNCHSCVYDLDDKCPRAEQRASPLTSVAVVGL 662
DB 598 GIMGENNTL-VWKYDANNVCHLCANCTYGCAGPGLQCEVWPSGPKIPSIATGVGL 656
QY 663 LVVVLGVVFGI-LIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAOMBLKETEIR 721
DB 657 LFTIV-VALGIGLFRKRRIIVKRTIRLLQRELVPLTPSGEAPNOAHLLKETERK 715
QY 722 KYKVLGSGAFGNNFTVSWLBPVKVASHLEV---LENTSPRANKIIDEAYVMAGV 777
DB 716 KIKVLGSGAFG---TVYKGLWIPGEKEKIVPAIKELKEATSPYANKIIDEAYVMASV 771
QY 778 GSPYVRLIGLCTSTVQVLTOLMPYGCILDHVRENGLSGODLLNMQAKMSYLE 837
DB 772 DNEHVCRLIGLCTSTVQVLTOLMPYGCILDHVRENGLSGODLLNMQAKMSYLE 831
QY 838 DVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKXMALESILR 897
DB 832 DRLVLRDLAARNVLYKTHQKHTIDFGAKLGLABEKYHAGKVPKXMALESILR 891
QY 898 RFTHQSDVWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPPICITDVMIMYK 957
DB 892 IYTHQSDVWSYGVTTWELMTFGSKPYDGIIPADISILKGRLLPQPPICITDVMIMYK 951
QY 958 CMMIDECPPRELVESESRMAPDQRFVVQV-NEDGAPAPLSTFPRSLLEDQMD 1016
DB 952 CMMIDSPKRELILESCMARPDQKTLVIGQGBKMLBPSTNSFRALMDEEDMD 1011
QY 1017 LVDAEEYLPQGGFFCPDPAPAGGVVHHRSSSTRSGGDLTLGLEPSEEARSPILA 1076
DB 1012 VVDADDEYLPQGGFF-----NSPT-----SRTPL 1037
QY 1077 PEEGASDVFPDDDLGMAAKGQSLPTHDPSPLOQYSEDPYPLPSET--DQVAPLTC 1134
DB 1038 SLSLATS---NSTVACINRNGSCRVEDAFLQRYSSDPTGAVTEDNIDFL----- 1087
QY 1135 POBEVYNQDPVRFPPSPREGLPAPAPGATLEBAKTLSPKNGNVVDFVFGAVENP 1194
DB 1088 PVPEVYNQ-SVKKRPAQSVQNVVHNQPLHP-----APRDLHYQN--PHSNAVGNP 1136
QY 1195 EYL-TPOGGAAPRPPAPFSPAPDNLTYWDQ-----DP-----PERGAPSTF 1237
DB 1137 EYLNIAQ-----PTCLSGFNSPALMICKQSHQSLNDPQODFFPFXETPNQIF 1187
QY 1238 KGTPTAENPEYLGNDVP 1254
DB 1188 KG-PTRENAEYLRVAP 1203

RESULT 6
TVCHLV
epidermal growth factor receptor precursor - chicken
N:contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C:species: Gallus gallus (chicken)
C:date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000
C:Accession: A27720, A00643

R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ulrich, A.; Vennart
Mol. Cell. Biol. 8, 1970-1978, 1988
A>Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mol
A:Reference number: A27720; MUID:88261272; PMID:3260329
A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M
Cell 41, 719-726, 1985
A>Title: C-erbB activation in ALV-induced erythroleukemia: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAY>
F:31-654/Domain: extracellular #status predicted <EXT>
F:61-307/Domain: EGF receptor extracellular domain repeat <EB1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EB2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136-202,280,361,370,422,575,580,615,635/binding site: carbohydrate (Thr) (covalent) #
F:992,650/binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 43.8%; Score 2983.5; DB 1; Length 1223;
Best Local Similarity 47.5%; Pred. No. 2,661,21;
Matches 619; Conservative 174; Mismatches 354; Indels 155; Gaps 29;

QY 8 RWGLLALLPFGAA-----STQVCTGTMDKLRPRAPETHLMDLRLHYQGGQVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKVCGGTNNKLTQGHVEDHFTSLQRMVNCCEVLSNLE 72
QY 62 LTLPTNALSIFQDIOEQVQYLLAHNQVQYRPLRLALVNGTQLEFEDVNLAYLNDG 121
DB 73 IYVEHNRDLTFLKTLQEVAGYVLLALNVDVLPENLDIIRGNLYLNDSPALAYLSNH 132
QY 122 PLNNTPTVAGSPGGLRELQSLSTELKGVLIQRNPQCYODTLIMKOIPIKNNQAL 181
DB 133 -NNKTO-----GIRELPKMRSLIILNGYKISNNPCLCMNDVLLMNDIIDTSRK-PL 182
QY 182 TLID-TNRRBACHPCSPKGRSGRCESESDQSLTRVCAAGCA-RCKGRLPTDCHEQ 239
DB 183 TVLDFASNLSSCKCKCHPNCCTEDHCWAGQNCQTLTKVCAQCCSGRCRGKVPSPCHNQ 242
QY 240 CAAGCTGPRHSDCLACTLHNHSGICEHCPALVTYVTFEEMPNREGYTGASCTYAC 299
DB 243 CAGGCTGPRSDCLAKRKRFRDQCTKOTPRVLYNPITTYQMDVPEGKYSFGATCVBEC 302
QY 300 PNYISTDVSGCTLVCPHNOEYVADGTQRCSEKSKPCARCYGGLGMOYIRANKFTGI 359
DB 303 PHNYVYTDHSGCVRSQNTDYEV-EENGVRKCKKCGCLCSKONGIGIOLGILS-INA 360
QY 360 TELE-PAGCKKIFGSLAFIPESFDGPANNTAPLQPEQVETLEITGYIYICAMPDS 418
DB 361 TINDSRKCTKINGDVSLPVAFLGAPFKTLPDPKLDVRRVTKESIGSFLLIAMDPND 420
QY 419 LPDLASYFQNLQVIRGLIHNGAYSLTQGLIGISWGLSRSLRSLGSLALIHNTLCFVH 478
DB 421 APTDLVAFENLEIRGTQKHQGYSLAVVNLKISGLRSLKISQDIALIMKRLCTAD 480
QY 479 TVPWQDLFRNPQALHTANRBEDECVEGLACHQICARGHCWGGPRTQCVNCSQFLRQG 538

DB 481 TNWRSPLATOSQKTKIQLNKNKDCTDRHVCPDLGSDVCMGPGPHCSRCFFRSKX 540
QY 539 ECVESCRVLQGLPRFYVYVARNCLPCHPCCQONG----SVTPGPEADQCVACAHKDPDF 595
DB 541 ECVKQCNLLQGEPRFERDSKCLPCHSCLQNSSTAYVTTQSGPDPDCKMCAHIDDPH 600
QY 596 CVARCPSPGKPDLSMPYMKFPDEBGAQCPPICTHSCVOLDLDDKGCBAERASPLTSIV 655
DB 601 CVKACPAVLGENDTL-VKXADANAVQGLCHPCTROCKPGLGEGC---NGSTPBLA 656
QY 656 SAVV-GILLVVLGVFGILIKRQOKIRKYMRLLOETELVEPLTSGAMPQAMRI 714
DB 657 AGVVGGLCLVVGIGIGILYLRH-HYRKKTLRLLOERELVEPLTSGAPQAHRI 715
QY 715 LKETEELRVKVLGSGAFGNNFTYSFWLRYVKVASHEV---LRETSKANKELLDE 770
DB 716 LKETEELRVKVLGSGAFG---TYKGLWPEGERVKIPIVAKELRETSKANKELLDE 771
QY 771 AYVAGVSPYVSRLLGICLTSTVQLVTLQMPYGLDHYENRGRGLSGDILLWCMQIA 830
DB 772 AYVAVSVNPRVCRLLGICLTSTVQLVTLQMPYGLDHYENRGRGLSGDILLWCMQIA 831
QY 831 KQMSYLEVRLVHRDLAARVLYVSPNHVKTITDGLARLDIDETEHADGKVPYKMA 890
DB 832 KQMYLERRRVHRDLAARVLYVTPQHVKTITDGLAKLIGADEKEYAEGKVPYKMA 891
QY 891 LESIRRRFTQSDVSGVTVWELMPGAKPYGIPARETLDLEKGERLPORPCTID 950
DB 892 LESIRRLYTHQSDVSGVTVWELMPGAKPYGIPARETLDLEKGERLPORPCTID 951
QY 951 VYIMVYKCMWIDSECRPRELVEFSFARMDFORFVYIQ-NEDLGPAUSDSTFYRLL 1009
DB 952 VYIMVYKCMWIDSECRPRELVEFSFARMDFORFVYIQ-NEDLGPAUSDSTFYRLL 1011
QY 1010 EDDMGDLVDAEYLYVPOQGFCCDPAAGACGMVHHRSSSTSGGDLTLGLPSEEE 1069
DB 1012 EEDMEDIVDAEYLYVPOQGF-----NSPST----- 1038
QY 1070 APRSPL-----APSEGAQDVFDGDLGGAAGLQSLPTHPSPLOYSEDPVLPSET 1124
DB 1039 -SRPDLSSLSATNSNTNID-----RNQGHVYRDSFVQYSSPTGNFLES 1089
QY 1125 --DGVVAPLTSPOPEYVQDPVRPQPSREGPLAARPAATLERAKTLPKNGVYK 1182
DB 1090 IDDGFL-----PAPEVYNQ--LMPKPS-----TAMQONQLYN 1120
QY 1183 DVF-----AFGAVNEEYLLPQGAAPQHPHAPSAFONLYVYQD-- 1225
DB 1121 NISLTAISKLPMSDRYQNSHTAVDNEEYL-----NTQSPPLAKTVFSSSPWYIQSG 1172
QY 1226 -----DPE-----RGAPSTFKGTPTAENPEYGLDVP 1254
DB 1173 NHQNLNDPVOYQDFLPNETKPNGLLAKVPAENPEYLYVAVP 1214

RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlon, G.W.; Foy, L
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A>Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide
A:Reference number: A47253; MUID:93189574; PMID:8393326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <FLO>
A:Cross-references: GB:L07868; NID:G337359; PIDN:AA859446.1; PID:G337360
A>Note: sequence extracted from NCBI backbone (NCBI:P126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology


```

Db      63 LETTYQENODLSFLOSIOEVGYU IANNEVSTIPVLNRLIRGNLYEGNFTLLWNS 122
QY      120 GPELNTTPTVAGSPGELRELOQLSLTEILKGVLIQIORNPOLCYODTILMKDIFHKNOL 179
Db      123 YCK-NPSSP--DYVUGLQKQLOLSNLTEILSGVAKSHNPLCNVETIWMWIDVKTSP 179
QY      180 ALTLIDTNSRAHPCSPMKSGRCGSESSDCCSLTRIVCAGC-ARCKGRLPTDCHE 238
Db      180 TMLIPPAFERQCKDHGCVNGSCWAPGPHGCKFTKLCAEQCNRRGRGKPIDCCNE 239
QY      239 QCAAGCTGPAHSDCLACLFHNHSGICELCPALVTYNTDTEFSMPREGRYFGASCVA 298
Db      240 HCGAGCTGPRATDCLACRPFNDGTCTDCTCPKXIYDIHQVNDPNKITYFGACVGE 299
QY      299 CPYNYLSTVGSCTLVCPHNDVTAEDGTQRCCKSPQCARVCGLGMOYIKANSKFTG 358
Db      300 CPENYVATE-GACVRCGSAQMLEVD-ENKRSCKPDCGCPKVCIDIGISL-SNTIAVN 356
QY      359 ITFL-EPAGCKIFGSLAFLPESFGDDPASNTAPLQPELOVFEETEEITGYLYISAWPD 417
Db      357 STIRSFNSCTKINGIILNRNSFEEDPHYKIGTMDPELMLNLTIVKELTGYLVIMWPE 416
QY      418 SLFDLSVFNLOVIRGRILHNGAYS-LTLQGLISWLGSLRSLGSLALIHNTLFCF 476
Db      417 NMTSLSVFQNLTEIRGRTFSRGFSFVAVQVHQLQWLGRSLKEVAGAVILKNTLQRLY 476
QY      477 VHTVPMDQLFNPHQALLTANRPEDECVGELAGHQLCARHGCNPGPTQVNGSQIFR 536
Db      477 ANTIMNRRLFRSDQIEYDART-----ENQTCNNESESDGCW-PGPTVCSGLHYDR 528
QY      537 GQCEVECEKVLQGLPREYVNAHRLCPHCECPQNSVYCFGEADQCAVACHYADPPC 596
Db      529 GRCVASCNLLQGEPEADQVDCVQCHQELVQDLSLTCTYGRPNACSKSAHFQDGPCC 588
QY      597 VACPSGVKPLDSYPMIKFPDEEGACQPPINCHSCVLDLDDKGPABORASPLTISVS 656
Db      589 IPRCPHGLDGDPTL-IMKYADKMGCCQPCQONTCQGGSGFELSGRGD-IVSHSLAVG 646
QY      657 AVVGLLVVVLGVFGILKRBQOKIRKYMRLIOETLVEPLTPSGMAMQAKRIK 716
Db      647 LVSGLLITIVIALIVLRLRRRIK-RKSTIRCLQEKELVPLTPSGAPQAPLRIK 705
QY      717 ETELRKYVLAGSAGFNNFTV-SFW-----LRVYKASASHELENTSPKANKELL 768
Db      706 ETEFKKDRVLGSAFG---TYXKGLMNDGENIRPVA-----IKVLRATSPKYNQDEL 757
QY      769 DEAYVWAGVSPYVSLGICLTSTYQVLTQMLPFQCLLDHYRNRGRIGSGDILNMCQ 828
Db      758 DEAYVWASVDHPHVCRLGICLTSAVQVLTQMLPYGCLLDYRQHERICQGLNLMVCQ 817
QY      829 IAKGMSYLEDVLAHEDLAARVLYKSPHVKITDFGLARLIDIDETEHADGKVPYIKM 888
Db      818 IAKGMVYLEERHLYHRLAARVLLKNPHVKITDFGLSKLLTAEBKEVQADGKVPYIKM 877
QY      889 MALESTLRRFRTHQSCVWSYVTTWELMTFGAKPYDGI-PAREIPDLLEKESRLPPRICT 948
Db      878 MALESTLQWTTYHOSVWSYVTTWELMTFGSKPYDGI-PAKEIASVLENGESLPPRICT 937
QY      949 IIVYIMVCKMIDSCPRPRFELVSESRMARDDQRFVIONEDLGASPLDSFFYSL 1008
Db      938 IEVYIILCKMIDBSRPRFRELJGFSQMKADSRVIVIG---NLPSLDRPLFRL 994
QY      1009 LEDDDMGDLVDAEYLVPQGFCCPDPAAGAGVWHRRSSSTRSGGDLTLGLPSEE 1068
Db      995 LSSDD--DYVDADEYLLPYKRI-----NKGGS----- 1019
QY      1069 EAPRSPLASBEGAGVDFDGLGMGAQKQSLPHTDPSPLQRYEDPTV-PLBSETGY 1127
Db      1020 ----EPCLEPTGH-----PVRENSITLRNISTDPLQNALEXDLGH 1055
QY      1128 VAPLTCSPQPEYVNOPVARPC-----PSPRE-----GPLP-AAPAGATLTERAKTIS 1174
Db      1056 -----EYVNOGSETSRSLSDIYNPNVEDLTDGNGPVSJSSQEAETNFSRPEYLN 1105

```

```

QY      1175 PKNGVNDVFAAGAVENPEYLTPOGGAAPQHPHPAFSPAFDNLVYWDODPBERGAP 1234
Db      1106 TNQNSL---PLVSSGSMDDPDY---QAG-----YQAAF-----LQGTGALT 1140
QY      1235 STFKCTPFAENPEYLG 1250
Db      1141 GNGMFLPAENLELYLG 1156

RESULT 9
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C/Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C/Accession: A36223; 159164
R/Kruse, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A/Title: Isolation and characterization of ERB3, a third member of the ERBB/epider
A/Reference number: A36223; MUID:90083234; PMID:26875
A/Accession: A36223
A/Status: Preliminary
A/Molecule type: mRNA
A/Residues: 1-1342 <XPA>
A/Cross-references: GB:M29356
R/Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A/Title: Molecular cloning and expression of another epidermal growth factor recept
A/Reference number: 159164; MUID:90311312; PMID:2164210
A/Accession: 159164
A/Status: preliminary; translated from GE/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A/Cross-references: GB:M34309; NID:G183990; PIDR:AAA35979.1; PID:G306841
C/Genetics:
A/Genes: GDB:ERB3; HER3
A/Cross-references: GDB:119880; OMIM:190151
A/Map position: 12q13-12q13
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C/Keywords: ATP; phosphotransferase
F/707-972/Domain: protein kinase homology <XIN>
F/715-723/Region: protein kinase ATP-binding motif

Query Match      33.9%; Score 2307.5; DB 2; Length 1342;
Best local similarity 39.4%; Pred. No. 3.5e-92;
Matches 519; Conservative 196; Mismatches 460; Indels 143; Gaps 36;

QY      10 GLIALALPPGAA--STQVCTGTDMKRLRPASPEITHLMLRHLVYCGVQGNLELYLPT 67
Db      11 GLTFLARSGSEVNSGAVCCGTNLGLSVTDGAENQYQTLVLYRCEVWGNLELYLGH 70
QY      68 NASLSFIQDIOEVGYVLIHNVQVQVPLQRLIRVGTQLFEDNYALAVALONGDPLNNT 127
Db      71 NADLSFIQWREVGYVLIHNVNNESTPLPLNLRVVRGTQYVYGFAIFVM-----LNYNT 125
QY      128 PVTGASPOGARELOLRSLTEILKGVLIQIORNPOLCYODTILMKDIFHKNQGLALTLIDTN 187
Db      126 ----NSSHALRQLRLTLEILSGVYIEKDKLCHNDITLDMKRIYVDRD---AEIVKD 178
QY      188 RSRACHPCSPMKSGRCGSESSDCCSLTRIVCAGC-ARCKGRLPTDCHEQCAAGCTG 246
Db      179 NGRSCPPCHEYCKG-RCMGPGSEDCQTLTITICAPQCNHGCFGNPNQCHDEAGCSG 237
QY      247 PKHSDCLALHFNHSGICELCPALVTYNTDTEFSMNPGRITFGASCYTACRYNYLST 306
Db      238 PDDTCFACRHFNDGSAVPCQPLVYNLTQLEENPHTKYQYGGVCAVASCPEHNV-V 256
QY      307 DVGSCTLVCPHNDVTAEDGTQRCCKSPCARVCGLGMOYIKANSKF--IGITFL- 363
Db      297 DQSCVACGPPDKAEVD-KYGLMKCEPCGGLCPKACGCTG-----SSRQGYDSSNIDG 350
QY      364 FAGCKIFGSLAFLPESFGDDPASNTAPLQPELOVFEETEEITGYLYISAWPDLDS 423

```


Db 351 FVACTKILGNLDELITGLNGDPWPKI:PALDPEKLVNFRVREITGYLNIQSWPBMHNF 410
 Qy 424 VFNQLOVIRGRILHNGAVS-LTLOGIGISWLSRSRLREGSGALAHNTHLCFYATVPM 482
 Db 411 VFNLTITIGRSLYNGRFSLLIMKNLVNLSLGRSLKEISAGIYISANRQLCYHSLWM 470
 Qy 483 DQLEFRPHQALHTA-NRPEDECVGEGGLACHQICARHCGWPGPTOCVNCQPLRGQECV 541
 Db 471 TKVLGRPTERLDIKHNRPRRDCVAEGKVCPLSSCGSGWPGQCLSCBNRSRGVCV 530
 Qy 542 EECRVLOGLPREYVNAHCLPCHPEOCPONGSVTCFPEADQCAVACHYKDPFCVACBP 601
 Db 531 THCNFLNGEPREPEHAECEFSCHPEOCPEGTATCGSSGSDTCAQAHFRDGHCVSSCP 590
 Qy 602 SGVCPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDDDDKCPAECRA---SPITSIWA 657
 Db 591 HGVLG--AKGPITYKYPVQNECRPHCHNCQSGKPELODCLQGLTILGKTHLTALIV 648
 Qy 658 VVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNOAQRILK 716
 Db 649 IAG--LVVIFPMLGTFLLWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVLARIFK 705
 Qy 717 ETELKRVKVLGSGAFNNFTVSMFLRVPKVSASH---LEVIRENTSPKANKELIDEAY 772
 Db 706 ETELKRVKVLGSGVFG---TVKGVWIPGESIKIPVCIKVIEDSGKSGFOAVTDHML 761
 Qy 773 VMAVGSPIVSRLLIGLTSTVQLTQMLPYGCLLDHVRENRGLSODLLWMCQIAGK 832
 Db 762 AIGSLDHAHIVRLICPGSSQLQVLTQYPLGSLDLHVQRGALPQLLLMWGVGIAGK 821
 Qy 833 MSYLEDVRLVHRPLAARNVLKSPVNHVKTIDPGLALLDDDETEHADGKVPKIMMALE 892
 Db 822 MYLLEHGMVHRLAARNVLKSPVQVADFGVADLPPDDQLLYSEAKPIKMALE 881
 Qy 893 SILRRFTHOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPITCIDY 952
 Db 882 SIHFQKVTQSDVWSYGVTVWELMTFGAEFYAGLRLAEVDDLKEGERLAPQICITIDY 941
 Qy 953 MIVKCMMDISECRPRELVSFESKMARDPORFVYITQNEGLPA---SPLDSTFYRSLL 1009
 Db 942 MIVKCMMDISECRPRELVSFESKMARDPORFVYITQNEGLPA---SPLDSTFYRSLL 1000
 Qy 1010 EDDMDGLVDAEEYLVQCGFPCPDPAFGAGWVHHRSSSTRSGGDLTGLDEP-SEE 1068
 Db 1001 EYELFEFLDLDLDELEED-----NLATTIGSALSLPVGTLNRP 1041
 Qy 1069 EABRSLPABEGAGSDVFDGLMGAKGIQSLPTH-DSPLOQYSEDPVPLP----- 1121
 Db 1042 RGSQSLSPSSGY-MPMNQNIGESCOESAVSGSSERCPVSLH-----PMRGLAS 1094
 Qy 1122 SETDGYVA-----PLTCSPOPE---YNNQPDVRQPPSPREGP----- 1156
 Db 1095 ESSEGHVTSGEALQEKVSMCRSRSSRPRRGDAHYHSQRHSLTPTPLSPPELEEE 1154
 Qy 1157 -----LPAARPAATLERAKTLP-SKQNGV-----KDVAFGAVENPEYLPQOGA 1203
 Db 1155 DVNGYVMPDTHLKGTPSSREGTLLSVGLSLCTEBEDD-----EYEVNMRRRH 1206
 Qy 1204 AQPHPRPAPSPAFNDLYWD-----QDPRKGAAPPSTKGPPTAENPYL 1249
 Db 1207 SP-PHPRPSSLEELGYEYMDVGSLSASLGSTQSCPLHPVIMPTAGTTPDEDEYEM 1263

RESULT 10

JC4387

epidermal growth factor receptor homolog precursor - rat

N.Altimate names: ErbB3 protein; HER3 protein

C.Species: Rattus norvegicus (Norway rat)

C.Date: 17-Jan-1996 #sequence, revision 19-Apr-1996 #text, change 13-Nov-1998

C.Accession: JC4387

R/Heller, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A>Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A:Reference number: JC4387; MUID:96096535; PMID:8522190
 A:Accession: JC4387
 A:Molecule type: mRNA
 A:Residues: 1-1339 <HEL>
 A:Cross-references: GB:U29339; NID:9915389; PID:9915390
 A:Experimental source: Liver
 A:Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res
 C:Comment: This protein is a functional heregulin receptor that transduces signals
 A:Genes: ErbB3
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
 C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote
 F:1-19/Domains: signal sequence #status predicted <SIG>
 F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
 F:40-659/Domains: transmembrane #status predicted <TMM>
 F:705-970/Domains: protein kinase homology <XIN>
 F:713-721/Region: protein kinase ATP-binding motif
 F:939,1051,1156,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Ty

Query Match

32.6%; Score 2221.5; DB 2; Length 1339;

Best Local Similarity 39.5%; Pred. No. 1.8e-88;

Matches 509; Conservative 173; Mismatches 440; Indels 167; Gaps 37;

Qy 3 LAALCRWGLLALPFGCA---STQVCTGDMKRLPASPETHLDMRLHYQCGQVQGN 59
 Db 7 LQVLC---FLLSIARSGSENGNSQAVCPGTINGSLVGDADNOYQTLTKLYEKEVWGN 62
 Qy 60 LELTLPFNASISFLDIOEYQVYLLAHNOVROVPLORLIRIVGTOLFEDVNAVLVDN 119
 Db 63 LEIVTGNADSLFLQWIREVTAYLVVAMEFSLPLNKLNVAGTQYDQKFAIFVM-- 120
 Qy 120 GDPLNTTPTVGTASFGGLRELQSLTEILKGVLIQNPOLCYQDTLLMKDIFHKNNQL 179
 Db 121 ---LNVNT---NSSHALRQLKFTQLEILISGVYIEKNDLCHMDITDMDIVR-- 170
 Qy 180 ALTLIDTNRSAACHPCSMCKSGKCGWSSSDCSLRTVTAGGC-ARCKELPTDQHE 238
 Db 171 GAELIVKKNAGNCPCEHEVCCKG-RCKWGGPDDCQILTKTICAPQCNCGCFGPNPQOCHD 229
 Qy 229 QCAAGCTGPKSDCLACLFHNSGICELHCPALTYNTDTEESMPNREGRTTFASCVTA 298
 Db 230 ECAGSGSPQDIDCFACRFNDSGACVRCGEPLVYNKLTQLLEPNHTKQYGVCVAS 289
 Qy 239 CPYNYLSTDVSGCTLVCPILHNOEYTAEDGTORCEKSKPCARVYGL--GMQYIKANSKF 356
 Db 230 CPNHFV-VDOYFCVACAPCPDMEVD-KIGLKPCPCGGLCEKAGEGSGSGSRQYQTVSSN 347
 Qy 357 IGITELFAGCKKIFGSLAFIPESFDDGPASNTAPLOEOLQVETLEIIGYLYISAWP 416
 Db 348 ID---GFVNCITKILGNLDELITGLNDPWHKIPALDPEKLVNFRVREITGYLNIQSWP 403
 Qy 417 DSLPDLVFNQLQVIRGRILHNGAVS-LTLOGIGISWLSRLSELGSLALHNNTHLC 475
 Db 404 PHMNFVSFNSLITIGRSLYNGRFSLLIMKNLVNLSLGRSLKEISAGIYISANQOLC 463
 Qy 476 FVHTVPMQULFRNHQALHTA-NRPEDECVGEGGLACHQICARHCGWPGPTOCVNCQSF 534
 Db 464 YHSLNMTRLLRGSEERLDIKYPRGECCLABEKVCDPLSSCGSGWPGQCLSCBNRSRGVCV 523
 Qy 535 LRQGEVCECRVLQGLPREYVNAHCLPCHPEOCPONGSVTCFPEADQCAVACHYKDP 594
 Db 524 SREGCVTHCNFLNGEPREPEHAECEFSCHPEOCPEGTATCGSSGSDTCAQAHFRDGH 583
 Qy 595 FCVACRSGVAPDLSTYMPIMKFPDEEGACOPCPINCTHSC-VLLDKGCPAEGASPLT 652
 Db 584 HCVASCPGILG--AKGPITYKYPVQNECRPHCHNCQSGKPELODCLQGLAEVLMKRFH 641
 Qy 653 SIYAVVGIILVVLGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNOAQ 711
 Db 642 LVIAVTVG--LAVILMLIGSFLVWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVL 698
 Qy 712 WRILKTELKRVKVLGSGAFNNFTVSMFLRVPKVSASH---LEVIRENTSPKANKEL 767

Db 699 ARIFKETELEKLVGSGVGF---TVHKGIWIPESGSIKIPVCIKYIEKSRGSGFQAV 754
 Qy 768 LDEAYVAVAGVSPVSSLLIGLITSTVQVLTQIMPGCLLDHYRENRGRGLSGDILLNMGK 827
 Db 755 TDHMLAVGSLDHAIVTLGLICPGSSIQVLTQVLTPLGSLLDHYKORRETLPGQLLNMGV 814
 Qy 828 QIAKGMVSYLDEVDLVNHDLAARVLYKSPNHVKITDGLARLLDIDETEVHADGKVPFK 887
 Db 815 QIAKGMVSYLDEHSHVNRDLKRVNMLKSPQVAVDPGVADLLPPDDKQLHSEAKTPFK 874
 Qy 888 WMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPIIC 947
 Db 875 WMALESIFGKYTHQSDVMSYGVTVWELMTFGAEPAAGLAEIPDLLEKGERLPQPIIC 934
 Qy 948 TIDVYVIMVKCMKIMIDSECRPRRELVSERMAPRDPQRFVIVIONEDLGASPIIDSTFYRS 1007
 Db 935 TIDVYVIMVKCMKIMIDENIRPTFKELANEFTRMARDEPRVLYIKRAS-GEQTP--PAEBS 991
 Qy 1008 LLEDDMDGLVDAEEYLVPOQGFPCPDPAAGAGVWHHRSSSTRSGGDLTLGLPBB 1067
 Db 992 VLTTEL---GEAELEPEL-----DLDLDEABE 1017
 Qy 1068 E-----EAPRSPPLASEG-----AGSDVPDGLGMAKGLQ 1099
 Db 1018 EGLATSLGSLSLPTGTLTPRGSGSLSPSSGYMPMNQSLGEACLDSAVLGGREGFSR 1077
 Qy 1100 SLTTPSPLOQRSEDPVLPSETDGV---APL-----TC-----SPQPE---Y 1139
 Db 1078 PISLH-PIDRGR-----PASESSGHVTGSFAELIOEKVSVCRSSRSRSPRFGDSAY 1129
 Qy 1140 VNQPDVPRPPSPREGR-----LPAARPAATLERAKTLSP-GKNGV----- 1181
 Db 1130 HSGRHSLLTPVTPLSPGLLEBDGNGVMPDTHLRASSSRECTLLSVGLSSVLTGREED 1189
 Qy 1182 KDVFAGGAVENVEYLTPOGAPPOPHPP 1210
 Db 1190 ED-----EEYVMRRKRGSP-PRPP 1209

RESULT 11

TVEVIV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
 N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
 C:Species: avian leukosis virus, ALV
 C>Date: 31-Dec-1991 #sequence_revision 11-Dec-1991 #text_change 11-Jun-1999
 C:Accession: B00643; A00643
 R:Hilgen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rotman, F.M.; Citterenden, L.B.; Raines, M.
 Cell 41, 719-726, 1985
 A>Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
 A:Reference number: A00643; MUID:85228222; PMID:2988784
 A:Accession: B00643
 A:Molecule type: mRNA
 A:Residues: 1-698 <NLS>
 A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PIDN:AAA48763.1; PID:g211750
 A:Note: in Genbank entry, CKBEBB, release 109.0, the source is designated as Gallus gal
 C:Comment: This protein is synthesized as a gag-env-erbB protein.
 C:Genetics:
 A:Gene: gag-env-erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
 F:1-6/Product: gag protein (fragment) #status predicted <GAG>
 F:7-56/Product: env protein (fragment) #status predicted <ENV>
 F:60-698/Product: protein-tyrosine kinase #status predicted <ERB>
 F:194-459/Domain: protein kinase homology <KIN>
 F:202-210/Region: protein kinase ATP-binding motif
 F:229/Active site: Lys #status predicted

Query Match

Best Local Similarity 24.3%; Score 1655.5; DB 1; Length 698;
 Matches 361; Conservative 81; Mismatches 145; Indels 133; Gaps 20;

578 GPEADOCVCAAHYKDPFCVACPSGVKEDLSYMPIMKPPDEGACOPPICHTSCYDUL 637

Db 60 GP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VKKYADANAVCOLCHPNCTRGCKGP 116
 Qy 638 DDKGCPAQRASPLTSIYSAVY-GILLVVLGVGVGILIKRRQKIRKYTMRLDQTEL 696
 Db 117 GLEGGP---NSGKTSIAGAVGGLCLVWVGLGLVLR--HIVRKTLRRLQDEEL 172
 Qy 697 VEPLTPSGAMPNOAMRLKETELRKVKVLSGAFGNFTVSVFLRVKVSASHLEY-- 754
 Db 173 VEPLTPSGAMPNOAMRLKETELRKVKVLSGAFGNFTVSVFLRVKVSASHLEY-- 754
 Qy 755 --LRENTSPKANKELDEAYVAVAGVSPVSSLLIGLITSTVQVLTQIMPGCLLDHYRE 812
 Db 229 KELRATSPKANKELDEAYVAVAGVSPVSSLLIGLITSTVQVLTQIMPGCLLDHYRE 288
 Qy 813 NRGRLSGDILLNMGKQIAKGMVSYLDEVDLVNHDLAARVLYKSPNHVKITDGLARLLDI 872
 Db 289 HNDNIGSQYLLNMGKQIAKGMVSYLDEVDLVNHDLAARVLYKSPNHVKITDGLARLLDI 348
 Qy 873 DETEYHADGKVPFKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIP 932
 Db 349 DEKEYHADGKVPFKWMALESILRRRFTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREIS 408
 Qy 933 DLLEKGERLPQPICTIDVYVIMVKCMKIMIDSECRPRRELVSERMAPRDPQRFVIVION 991
 Db 409 SVLEKGERLPQPICTIDVYVIMVKCMKIMIDSECRPRRELVSERMAPRDPQRFVIVION 468
 Qy 992 EDLGASPIIDSTFYRSLEDDMDGLVDAEEYLVPOQGFPCPDPAAGAGVWHHRSSS 1051
 Db 469 ERMLHPSPTDSKFYVTLMEEDMEDIVDAEYLVHQEFF-----NSPS 512
 Qy 1052 TRSGGDLTLGLPBBEERAPRSP-----ASEGASVVFQDGLGMAKGLQSLPTIDP 1106
 Db 513 T-----SRPLSSISATSNNSATNCID-----RNGGHPVED 546
 Qy 1107 SPLQRYSEDPVLPSET--DGYVAPLTCSPPQEVYVNPDPVPPSPREGRPLPARAPAG 1164
 Db 547 SFVQRYSSDPDIGNFLBESIDDGFL-----PAPEYVQ--LMPKKPS----- 585
 Qy 1165 ATLERAKTLSPGKNGVQVDF-----AFGAVENVEYLTPOGAPPOPHPP 1209
 Db 586 -----TAMVQNCIYNNISLTAISKLPMSRYQNSHSTAVDNEYL-----NTNQ 629
 Qy 1210 PPATSPAFDNLVYMDQ-----DPEP-----RGAPPTFGTTPAENPEYLGIDVP 1254
 Db 630 SPLAKTVFESSPTMIQSGNHQINIDNPYQODFLPNETKPNGLLKVPAAENPEYLVRAAP 689

RESULT 12

TYVUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain 1
 C:Species: avian erythroblastosis virus
 C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
 C:Accession: A00644; A38022
 R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
 Cell 35, 71-78, 1983
 A>Title: The erbB gene of avian erythroblastosis virus is a member of the src gene f
 A:Reference number: A00644; MUID:84026539; PMID:6313229
 A:Accession: A00644
 A:Molecule type: DNA
 A:Residues: 1-604 <YAM>
 A:Cross-references: GB:X01216; NID:g209676; PIDN:AAA42400.1; PID:g209678
 R:Debutre, B.; Henry, C.; Benabissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Mart
 Science 224, 1456-1459, 1984
 A>Title: Sequencing the erbB gene of avian erythroblastosis virus reveals a new type
 A:Reference number: A38022; MUID:84223957; PMID:6328658
 A:Accession: A38022
 A:Molecule type: DNA
 A:Residues: 1-28, 'W', '30-139', 'F', '141-145', 'V', '147-152 <DEB>
 A:Cross-references: GB:X02006
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specif

F:130-395/Domain: protein kinase homology <KIN>
 F:138-146/Region: protein kinase ATP-binding motif
 F:165/Active site: Lys #status predicted

Query Match

23.4%; Score 1592; DB 1; Length 604;
 Best Local Similarity 50.0%; Pred. No. 1e-61;
 Matches 347; Conservative 77; Mismatches 136; Indels 134; Gaps 18;

QY 587 CAHYKDPFCVAPCPGVPKPLSYMPIMKPFDEGACOPCPINCHSCVLDLDCGCAEQ 646
 DB 3 CAHFDGHCYKACAPAGVAGENDTL-VKYNADANAQOLCHPNCTRCCKRGLEGCC--- 58
 QY 647 RASPLTIVSAVY-GILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLPSGA 705
 DB 59 NGSKTPTSIAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRRLQERLEVEPLPSGE 117
 QY 706 MPNOAMPILKETELRKVKYLGSGAFGNNFTVSFMLRVPKVSASHLEV---LRENTSP 761
 DB 118 APNOAMHRLKETELRKVKYLGSGAFG---TLYKGLMIPGEEKVKYIPVAIKELREKTS 173
 QY 762 KANKELDEAYVMAVGSPPVSRLLGICLTSTVQLTQLMFYGCLLDHVENRGLSQD 821
 DB 174 KANKELDEAYVMAVSDNPVHCYCLLGICLTSTVQLTQLMFYGCLLDYIEHKONISQY 233
 QY 822 LLMNCMOIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDEGALDLIDETFYHAG 881
 DB 234 LLNMCVOIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDEGALDLIDETFYHAG 293
 QY 882 GKVPILKMALESILRRFTTHQSVDVSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERL 941
 DB 294 GKVPILKMALESILHRIYTHQSVDVSGVTVWELMTFGSKPYDGIIPASELSVLEKGERL 353
 QY 942 POPICITIDVYIMVWKCMIDSECRPRRELVSFSKMDPQCFVYIQ-NEDIGPASP 1000
 DB 354 POPICITIDVYIMVWKCMIDSECRPRRELVSFSKMDPQCFVYIQ-NEDIGPASP 413
 QY 1001 DSTFYRSLDEDDMGDLVDAEELVLPQGFPCPDPAAGAGVYHRRSSSTRGSGDLT 1060
 DB 414 DSKYRILMEEDMEDIVDAEELVLPQGF---NSFST----- 449
 QY 1061 LGLEPSEEARSPFL---APSEGAGDVFDDGLGMAKGLSLPTHPSPLORYSED 1115
 DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGGHPVREDSFYQRYSSD 491
 QY 1116 PTVPLPSET-DGVAPLTCPOPEYVNPQDPVPPSPREGPLPAARPAATLERAKTL 1173
 DB 492 PTVPLPSETIDGFL-----PAPEYVNPQ-LMPKPPSTAY----- 524
 QY 1174 SPGRNGVVKDVFAR-----OGAVENPEYLTPOGGAAPQHPPPAFS 1214
 DB 525 -----VONQIYNTLSLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNOSPLAK 570
 QY 1215 PAFDNLVYWDPPERKAPSTFKGTPTAENPEY 1248
 DB 571 TVFESSPYWISQNGH-----INLDNPDY 594

RESULT 13

S35745

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
 C:Species: avian erythroblastosis virus
 C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
 C:Accession: S35745
 R:Venustrom, B.
 submitted to the EMBL Data Library, March 1993
 A:Reference number: S35743
 A:Accession: S35745
 A:Molecule type: DNA
 A:Residues: 1-544 <VEN>
 A:Cross-references: EMBL:X12707
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-speci
 F:135-400/Domain: protein kinase homology <KIN>
 F:143-151/Region: protein kinase ATP-binding motif
 F:170/Active site: Lys #status predicted

Query Match

22.6%; Score 1541; DB 2; Length 544;
 Best Local Similarity 52.7%; Pred. No. 1.4e-59;
 Matches 333; Conservative 71; Mismatches 128; Indels 100; Gaps 17;

QY 578 GREADCCVCAHYKDPFCVAPCPGVPKPLSYMPIMKPFDEGACOPCPINCHSCVLD 637
 DB 1 GP-DHCKCAHFDGHCYKACAPAGVAGENDTL-VKYNADANAQOLCHPNCTRCCKRG 57
 QY 638 DKGCPAEORASPLTIVSAVY-GILLVVLGVVFGILIKRQOKIRKYMRLLOETEL 696
 DB 58 GLEGCP---NGSKTPTSIAGVVGGLCLVVGGLGIGLYLRR-HIVKRTLRRLQERLE 113
 QY 697 VEPLTPSGAMPQOAKRIKETELRKVKYLGSGAFGNNFTVSFMLRVPKVSASHLEV-- 754
 DB 114 VEPLTPSGAMPQOAKRIKETELRKVKYLGSGAFG---TLYKGLMIPGEEKVKYIPVA 169
 QY 755 --LRENTSPKANKELDEAYVMAVGSPPVSRLLGICLTSTVQLTQLMFYGCLLDHVE 812
 DB 170 KEREKTSKANKELDEAYVMAVSDNPVHCYCLLGICLTSTVQLTQLMFYGCLLDYIE 229
 QY 813 NGRIGSODLLNMCQIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDEGALDLID 872
 DB 230 HKDNIGSOYLLNMCVOIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDEGAL 289
 QY 873 DEFEYADGCKPIKMALESILRRFTTHQSVDVSGVTVWELMTFGAKPYDGIIPAREIP 932
 DB 290 DEFEYADGCKPIKMALESILHRIYTHQSVDVSGVTVWELMTFGSKPYDGIIPASELS 349
 QY 933 DLLEKGERLPOPICITIDVYIMVWKCMIDSECRPRRELVSFSKMDPQCFVYIQ-N 991
 DB 350 SVLEKGERLPOPICITIDVYIMVWKCMIDSECRPRRELVSFSKMDPQCFVYIQD 409
 QY 992 EDLGRASPJDSTFYSLDEDDMGDLVDAEELVLPQGFPCPDPAAGAGVYHRRSS 1051
 DB 410 ERMHLPSPDTSKRYRLMEEDMEDIVDAEELVLPQGF-----NSDS 453
 QY 1052 TRSGGADTLGLEPSEEARSPFL---APSEGAGDVFDDGLGMAKGLSLPTHP 1106
 DB 454 T-----SRTPLLSLSATSNNSATNCIDNG-----H-- 481
 QY 1107 SPLQRYSEDPTVPLPSETIDGFLVAPLTCPOPEYVNPQDPVPPSPREGPLPAARPA 1166
 DB 482 -----PVREDGFL-----PAPEYVNPQ-LMPKPPSTAYVONQIYNTLSL 519
 QY 1167 -LERAKTSPGRNGVVKDVFARFAGAVENPEYL 1197
 DB 520 AISKLPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14

S00727

kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi
 C:Species: avian erythroblastosis virus
 C:Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
 C:Accession: S00727
 R:Scotting, P.; Venustrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
 Oncogene Res. 1, 265-278, 1987
 A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus m
 A:Reference number: S00727; MUID:88217326; PMID:2897102
 A:Accession: S00727
 A:Molecule type: DNA
 A:Residues: 1-545 <SCO>
 A:Cross-references: EMBL:X06943
 C:Genetics:
 A:Gene: erbB
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; phosphotransferase
 F:135-400/Domain: protein kinase homology <KIN>

F:143-151/Region: protein kinase ATP-binding motif

Query Match 22.5%; Score 1534; DB 2; Length 545;
 Best Local Similarity 52.7%; Pred. No. 2.8e-59;
 Matches 333; Conservative 70; Mismatches 129; Indels 100; Gaps 17;

QY 578 GEPADQCACAHYKDPFCYACRPSGVKEDLSYMPWKPPDEGACQPCPINCTHSCVDL 637
 DB 1 GP--DHCMKCAHFIDGHCYKACPAVLGENDTL-VWKYADANAVCOLCHPNCRCCKGCP 57
 QY 638 DDGCPAEGQASPLTSIVSAVY-GILLVVLGVVFGILIRKROOKIRKYMRLLOETEL 696
 DB 58 GLEGCP--NGSTPSTIAAGVVGGLCLVVGIGLYLRR-HYVKRLRLRLQRELV 113
 QY 697 VEPDTPSGAMPNOAKRLIKETELRYKVLGSGAFGNFTVSPWLVPKVSASHLEV-- 754
 DB 114 VEPDTPSGAMPNOAKRLIKETELRYKVLGSGAFGNFTVSPWLVPKVSASHLEV-- 169
 QY 755 --LRENTSPKANEILDEAVYVAGVSPYVSLGICLTSTVQLTQMLPYGCLDHYRE 812
 DB 170 KELREATSPKANEILDEAVYVAGVSPYVSLGICLTSTVQLTQMLPYGCLDHYRE 229
 QY 813 NRGRLGSDLLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDI 872
 DB 230 HKDNIGSQVILNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDI 289
 QY 873 DETEYHADGSKVPIKMALESILRRFTHOSDVSYGVTYWEIMTGAKPYDGIAPREIP 932
 DB 290 DEKEYHAEGCKVPIKMALESILRRFTHOSDVSYGVTYWEIMTGAKPYDGIAPREIP 349
 QY 933 DLEKGERLPPICTIDVYMIWCMKMIIDSECRPRELSEFSRMAPDPORFVYIQ-N 991
 DB 350 SVLEKGERLPPICTIDVYMIWCMKMIIDSECRPRELSEFSRMAPDPORFVYIQ-N 409
 QY 992 EDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPPAPAGAGMYHRRSSS 1051
 DB 410 ERNHLPSPTDSKRYRTIMEBEDMEDIVDAEYLVPOQGF-----NSPS 453
 QY 1052 TRSGGCDLTLGLEPSEEARSP-----APSEGAGSDVFDGDLGMAKGLGSLPTHPD 1106
 DB 454 T-----SRTPLLSLSATSNNSATNCIDRNG-----H-- 481
 QY 1107 SPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYVNOQDVPPOPSPREGPLPARAPAGAT 1166
 DB 482 -----PVRDEDFL-----PAPRYVNO--LMPKRPSTAMVNOQIYISLIT 519
 QY 1167 -LERAKTLSPGKGVKVPFAPGAVENPEYL 1197
 DB 520 AISKLPMDSRYN-----SHSTAVDNPEYL 544

RESULT 15

B44776

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)

C/Species: avian erythroblastosis virus

C/Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 04-Feb-2000

C/Accession: B44776

R:Bruskin, A.; Jackson, J.; Bishop, J.M.; McCauley, D.J.; Schatzman, R.C.

A>Title: Six amino acids from the retroviral gene gag greatly enhance the transforming P

A/Reference number: A44776; M01D:9020603; PMID:1969616

A/Accession: B44776

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-540 <BRU>

A:Cross-references: GB:X52211

C:Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

F:130-395/Domain: protein kinase homology <KIN>

F:138-146/Region: protein kinase ATP-binding motif

Query Match 22.3%; Score 1517; DB 2; Length 540;
 Best Local Similarity 52.6%; Pred. No. 1.5e-58;
 Matches 328; Conservative 70; Mismatches 127; Indels 98; Gaps 16;

QY 567 CAHYDPPFCYACRPSGVKEDLSYMPWKPPDEGACQPCPINCTHSCVDLDDKCPAEG 646
 DB 3 CAHFIDGHCYKACPAVLGENDTL-VWKYADANAVCOLCHPNCRCCKGCPGLEGCP-- 58
 QY 647 RASPLTSIVSAVY-GILLVVLGVVFGILIRKROOKIRKYMRLLOETELVEPLTPSGA 705
 DB 59 NGSTPSTIAAGVVGGLCLVVGIGLYLRR-HYVKRLRLRLQRELV 117
 QY 706 MPNOAKRLIKETELRYKVLGSGAFGNFTVSPWLVPKVSASHLEV--LRENTSP 761
 DB 118 APNOAKRLIKETELRYKVLGSGAFGNFTVSPWLVPKVSASHLEV--LRENTSP 173
 QY 762 KANEILDEAVYVAGVSPYVSLGICLTSTVQLTQMLPYGCLDHYRENRGLSGD 821
 DB 174 KANEILDEAVYVAGVSPYVSLGICLTSTVQLTQMLPYGCLDHYRENRGLSGD 233
 QY 822 LLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDI 881
 DB 234 LLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLDI 293
 QY 882 GKVPKMALESILRRFTHOSDVSYGVTYWEIMTGAKPYDGIAPREIP 941
 DB 234 GKVPKMALESILRRFTHOSDVSYGVTYWEIMTGAKPYDGIAPREIP 353
 QY 942 POPPICTIDVYMIWCMKMIIDSECRPRELSEFSRMAPDPORFVYIQ-NEDGPA 1000
 DB 354 POPPICTIDVYMIWCMKMIIDSECRPRELSEFSRMAPDPORFVYIQ-NEDGPA 413
 QY 1001 DSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPPAPAGAGMYHRRSSS 1060
 DB 414 DSKFYRTIMEBEDMEDIVDAEYLVPOQGF-----NSPST 449
 QY 1061 LGLEPSEEARSP-----APSEGAGSDVFDGDLGMAKGLGSLPTHPD 1115
 DB 450 -----SRTPLLSLSATSNNSATNCIDRNG-----H-- 476
 QY 1116 PTVPLPSETDGYVAPLTCSPQPEYVNOQDVPPOPSPREGPLPARAPAGAT-LERAKT 1174
 DB 477 -----PVRDEDFL-----PAPRYVNO--LMPKRPSTAMVNOQIYISLIT 523
 QY 1175 PGKGVKVPFAPGAVENPEYL 1197
 DB 524 RYON-----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 05:28:12
 Job time : 30.2855 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 seconds

(without alignments)
4891.279 Million cell updates/sec

Title: SEQ4-710-730-14

Perfect score: 6813
Sequence: 1 MELALCRWGLLLALLPGCA.....TFKTPTAENPEYIGLDVAV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match Length	ID	Description
1	6602	96.9	1255	1 ERB2_HUMAN
2	5811	85.3	1257	1 ERB2_RAT
3	5795.5	85.1	1254	1 ERB2_MESAU
4	3033	44.5	1210	1 EGFR_HUMAN
5	3007	44.1	1210	1 EGFR_MOUSE
6	2852.5	41.9	1308	1 ERB4_HUMAN
7	2833	41.6	1308	1 ERB4_RAT
8	2575.5	37.8	1167	1 XMRK_XIPMA
9	2316.5	34.0	1342	1 ERB3_HUMAN
10	2244.5	32.9	1339	1 ERB3_RAT
11	1836	26.9	1426	1 EGFR_DROME
12	1638.5	24.0	634	1 ERBB_NLV
13	1595	23.4	703	1 EGFR_CHICK
14	1592	23.4	604	1 ERBB_AVIER
15	1524	22.4	540	1 ERBB_AVIER
16	1217	17.9	1323	1 LIT2_CAEEL
17	1142.5	16.8	245	1 ERB2_MOUSE
18	693	10.2	1363	1 ILPR_BRALA
19	656.5	9.6	1372	1 INSR_MOUSE
20	656	9.6	1300	1 IRR_MOUSE
21	655	9.6	1382	1 INSR_HUMAN
22	653	9.6	1383	1 INSR_RAT
23	647	9.5	1297	1 IRR_HUMAN
24	644.5	9.5	1300	1 IRR_CAVPO
25	632	9.3	1477	1 HTR7_AYDAT
26	630	9.2	1607	1 MTPR_LYNST
27	602	8.8	1367	1 IGR_HUMAN
28	584	8.6	1373	1 IGR_MOUSE
29	580.5	8.5	1370	1 IGR_RAT
30	580	8.5	1114	1 RET_HUMAN
31	579	8.5	987	1 EPB4_HUMAN
32	575	8.4	2146	1 INSR_DROME
33	569	8.4	1390	1 INSR_AEDAE

34	559	8.2	987	1 EPB4_MOUSE	P54761 mus musculus
35	551.5	8.1	984	1 EPB1_CHICK	Q07494 gallus galli
36	548.5	8.1	984	1 EPB1_RAT	P09759 rattus norv
37	544	8.0	902	1 EPB3_XENLA	Q91736 xenopus lae
38	542.5	8.0	984	1 EPB1_HUMAN	P54762 homo sapien
39	536.5	7.9	985	1 EPB4_XENLA	Q91571 xenopus lae
40	534.5	7.8	977	1 EPB2_MOUSE	Q03145 mus musculus
41	529.5	7.8	976	1 EPB2_HUMAN	P29317 homo sapien
42	525.5	7.7	985	1 EPB4_XENLA	Q91694 xenopus lae
43	525.5	7.7	1053	1 FAK1_CHICK	Q00944 gallus galli
44	524.5	7.7	986	1 EPB4_XENLA	Q91845 xenopus lae
45	523	7.7	1068	1 FAK1_XENLA	Q91738 xenopus lae

ALIGNMENTS

RESULT 1	
ERB2_HUMAN	STANDARD, PRT, 1255 AA.
ID ERB2_HUMAN	
AC P04626;	
DT 13-AUG-1987 (Rel. 05, Created)	
DR 13-AUG-1987 (Rel. 05, Last sequence update)	
DT 15-JUN-2002 (Rel. 41, Last annotation update)	
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)	
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell	
DE surface receptor HER2) (MLN 19).	
GN ERB2 OR HER2 OR NGL OR NEU.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX NCBI_Taxid=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=66118663; PubMed=3003577;	
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,	
RA Saito T., Toyoshima K.;	
RT "Similarity of protein encoded by the human c-erb-B-2 gene to	
RT epidermal growth factor receptor.";	
RL Nature 319:230-234(1986).	
RN [2]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=86070181; PubMed=299974;	
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;	
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the	
RT c-erbB-1/epidermal growth factor receptor gene and is amplified in a	
RT human salivary gland adenocarcinoma.";	
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).	
RN [3]	
RP SEQUENCE OF 737-1031 FROM N.A.	
RX MEDLINE=86016729; PubMed=2995867;	
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;	
RT "Characterization of a new allele of the human ERBB2 gene by allele-	
RT specific competition hybridization.";	
RL Genomics 15:426-429(1993).	
RL [4]	
RP VARIANTS VAL-654 AND VAL-655.	
RX MEDLINE=93194196; PubMed=8095488;	
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;	
RT "Characterization of a new allele of the human ERBB2 gene by allele-	
RT specific competition hybridization.";	
RL Genomics 15:426-429(1993).	
CC - FUNCTION. ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.	
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A	
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-	
CC ALPHA AND AMPHIREGULIN.	
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein	
CC tyrosine phosphate.	
CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS	
CC (POTENTIAL).	
CC - SUBCELLULAR LOCATION: Type I membrane protein.	

CC -1- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M11767; AAA35808.1; JOINED.
 DR EMBL: M11761; AAA35808.1; JOINED.
 DR EMBL: M11762; AAA35808.1; JOINED.
 DR EMBL: M11763; AAA35808.1; JOINED.
 DR EMBL: M11764; AAA35808.1; JOINED.
 DR EMBL: M11765; AAA35808.1; JOINED.
 DR EMBL: M11766; AAA35808.1; JOINED.
 DR EMBL: M11730; AAA75493.1; JOINED.
 DR EMBL: M12036; AAA35978.1; JOINED.
 DR EMBL: X03363; CAA27060.1; JOINED.
 DR PIR: A25491; A25491.
 DR PIR: A24571; A24571.
 DR HSP: P11362; IFGK.
 DR Genew: HGNC:3430; ERBB2.
 DR MIM: 164870; ERBB2.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR InterPro: IPR004019; YLP_motif.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP_2; domain; 1.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYKc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Polymorphism.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 653 675 POTENTIAL.
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 720 987 PROTEIN KINASE.
 FT BINDING 726 734 ATP (BY SIMILARITY).
 FT ACT_SITE 753 753 ATP (BY SIMILARITY).
 FT DISULFID 845 845 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 227 BY SIMILARITY.
 FT DISULFID 224 235 BY SIMILARITY.
 FT DISULFID 236 244 BY SIMILARITY.
 FT DISULFID 240 252 BY SIMILARITY.
 FT DISULFID 255 264 BY SIMILARITY.
 FT DISULFID 266 295 BY SIMILARITY.
 FT DISULFID 299 311 BY SIMILARITY.
 FT DISULFID 315 331 BY SIMILARITY.
 FT DISULFID 334 338 BY SIMILARITY.
 FT DISULFID 511 520 BY SIMILARITY.
 FT DISULFID 515 528 BY SIMILARITY.
 FT DISULFID 531 540 BY SIMILARITY.
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.
 FT DISULFID 567 584 BY SIMILARITY.
 FT DISULFID 587 596 BY SIMILARITY.
 FT DISULFID 600 623 BY SIMILARITY.
 FT DISULFID 626 634 BY SIMILARITY.
 FT DISULFID 630 642 BY SIMILARITY.
 FT MOD_RES 1139 1139 BY SIMILARITY.
 FT CARBOHYD 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARIANT 654 654 I -> V.
 FT VARIANT 655 655 /FTID=VAR_004077.
 FT VARIANT 655 655 I -> V.
 FT CONFLICT 1170 1170 /FTID=VAR_004078.
 FT CONFLICT 1255 AA; 137909 MW; 39E9DFDA04DC962 CRC64; P -> A (IN REF. 2).
 SQ SEQUENCE

Query Match Best Local Similarity 96.94; Score 6602; DB 1; Length 1255;
 Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELALCRWGLLALLPFGAASVTCVGTGDKMLRPAPEHMLDMLHLYGCGVQVGNL 60
 DB 1 MELALCRWGLLALLPFGAASVTCVGTGDKMLRPAPEHMLDMLHLYGCGVQVGNL 60
 QY 61 ELTLPPTNASLSPQDIQOEVOGYVLIHNVQVPLRLRIVRGTCLEFEDNYALAVLDNG 120
 DB 61 ELTLPPTNASLSPQDIQOEVOGYVLIHNVQVPLRLRIVRGTCLEFEDNYALAVLDNG 120
 QY 121 DPLNNTPTVTCASGGLRELQRLSLEILKGAVLIGNPOLCYDDTLIMKDIHPKNNOLA 180
 DB 121 DPLNNTPTVTCASGGLRELQRLSLEILKGAVLIGNPOLCYDDTLIMKDIHPKNNOLA 180
 QY 121 DPLNNTPTVTCASGGLRELQRLSLEILKGAVLIGNPOLCYDDTLIMKDIHPKNNOLA 180
 DB 121 DPLNNTPTVTCASGGLRELQRLSLEILKGAVLIGNPOLCYDDTLIMKDIHPKNNOLA 180
 QY 181 LTLIDNRSRACHPCSPMCKSRGCSSESDCGLRTVCAAGCARGKGPLPTDCHEOC 240
 DB 181 LTLIDNRSRACHPCSPMCKSRGCSSESDCGLRTVCAAGCARGKGPLPTDCHEOC 240
 QY 181 LTLIDNRSRACHPCSPMCKSRGCSSESDCGLRTVCAAGCARGKGPLPTDCHEOC 240
 DB 181 LTLIDNRSRACHPCSPMCKSRGCSSESDCGLRTVCAAGCARGKGPLPTDCHEOC 240
 QY 241 AAGCTGPHSDCLALCFHNSGICELCPALVTYNTDFSMPEDEGRYTFGASCVTACP 300
 DB 241 AAGCTGPHSDCLALCFHNSGICELCPALVTYNTDFSMPEDEGRYTFGASCVTACP 300
 QY 301 YNYLSTDVGSCTVLCPLHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
 DB 301 YNYLSTDVGSCTVLCPLHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
 QY 301 YNYLSTDVGSCTVLCPLHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
 DB 301 YNYLSTDVGSCTVLCPLHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
 QY 361 ELERAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIYSAMPDLP 420
 DB 361 ELERAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIYSAMPDLP 420
 QY 361 IOEPFAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIYSAMPDLP 420
 DB 361 IOEPFAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIYSAMPDLP 420
 QY 421 DLSVFQNLQVIRGRILNGAYSLTLQGLISWJLGRRLRELGSLALIHNNTHLCFHTV 480
 DB 421 DLSVFQNLQVIRGRILNGAYSLTLQGLISWJLGRRLRELGSLALIHNNTHLCFHTV 480
 QY 481 PMDGLFNNPQALHTANRDEDECVEGGLACHOLCARGHGMGHPPTQCVNCSQFLRQEC 540
 DB 481 PMDGLFNNPQALHTANRDEDECVEGGLACHOLCARGHGMGHPPTQCVNCSQFLRQEC 540
 QY 481 PMDGLFNNPQALHTANRDEDECVEGGLACHOLCARGHGMGHPPTQCVNCSQFLRQEC 540
 DB 481 PMDGLFNNPQALHTANRDEDECVEGGLACHOLCARGHGMGHPPTQCVNCSQFLRQEC 540
 QY 541 VEEBRVYQGLPREVYVNRHCLPCHPEQOPONGSVTCGPEADQCVACAHAKDPFVCAR 600
 DB 541 VEEBRVYQGLPREVYVNRHCLPCHPEQOPONGSVTCGPEADQCVACAHAKDPFVCAR 600
 QY 541 VEEBRVYQGLPREVYVNRHCLPCHPEQOPONGSVTCGPEADQCVACAHAKDPFVCAR 600
 DB 541 VEEBRVYQGLPREVYVNRHCLPCHPEQOPONGSVTCGPEADQCVACAHAKDPFVCAR 600
 QY 601 PSGVKPLSLYMPIMKPFDEGACQPCINCTHSCVDLDDKGCAPAEGRASPLTSIIAIVG 660
 DB 601 PSGVKPLSLYMPIMKPFDEGACQPCINCTHSCVDLDDKGCAPAEGRASPLTSIIAIVG 660
 QY 661 ILVVYVGVFGLIKRQOKIRKYTRRLLOETELVEPLTSGAMPNQAQMRILKXTEL 720
 DB 661 ILVVYVGVFGLIKRQOKIRKYTRRLLOETELVEPLTSGAMPNQAQMRILKXTEL 720

FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEO).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.3%; Score 581; DB 1; Length 1257;
 Best Local Similarity 85.3%; Pred. No. 6.8e-304;
 Matches 1077; Conservative 58; Mismatches 115; Indels 12; Gaps 5;

QY 1 MELAAALCRWGLLLALPPGAASCTGCTDMKRLPASPEHLDMLRHLYOGCCVQGNL 60
 DB 1 MELAAACWGFLLALPPGIAGCTGCTDMKRLPASPEHLDMLRHLYOGCCVQGNL 60
 QY 61 ELTYLPTNASLFLDIOEVOGVYLIANOVROVPLRLRYRGTOLFEEDNYALAVLDNG 120
 DB 61 ELTYLPTNASLFLDIOEVOGVYLIANOVROVPLRLRYRGTOLFEEDNYALAVLDNR 120
 QY 121 DPLNNTTPTVT-GASPGGLRELOLRSLTEILKGVYLIORNPOLCYQDTILWKDIFFHKNNQL 179
 DB 121 DPLNNTTPTVT-GASPGGLRELOLRSLTEILKGVYLIORNPOLCYQDTILWKDIFFHKNNQL 179
 QY 121 DPONNVAASTPGRPEGLRELQLRSLTEILKGVYLIORNPOLCYQDTILWKDIFFHKNNQL 180
 DB 121 DPONNVAASTPGRPEGLRELQLRSLTEILKGVYLIORNPOLCYQDTILWKDIFFHKNNQL 180
 QY 180 ALTLIDTRSRACHPSCSMCKSGSCWSESSSDCSLRTVCAGACACGKPLPDCCHEQ 239
 DB 180 ALTLIDTRSRACHPSCSMCKSGSCWSESSSDCSLRTVCAGACACGKPLPDCCHEQ 239
 QY 181 APVDIDTRSRACHPSCSMCKSGSCWSESSSDCSLRTVCAGACACGKPLPDCCHEQ 240
 DB 181 APVDIDTRSRACHPSCSMCKSGSCWSESSSDCSLRTVCAGACACGKPLPDCCHEQ 240
 QY 240 CAAGCTGPKHSDCLACIHFNSGICELHCPALVYNTDFESMNPREGRTFGASCYTAC 299
 DB 240 CAAGCTGPKHSDCLACIHFNSGICELHCPALVYNTDFESMNPREGRTFGASCYTAC 299
 QY 241 CAAGCTGPKHSDCLACIHFNSGICELHCPALVYNTDFESMNPREGRTFGASCYTAC 300
 DB 241 CAAGCTGPKHSDCLACIHFNSGICELHCPALVYNTDFESMNPREGRTFGASCYTAC 300
 QY 300 PYNVLTSDVGSCTLVCPLIHNOEVTAEEDGTORCEKSPCARVCGYGLMOYIKANSKEIGI 359
 DB 300 PYNVLTSDVGSCTLVCPLIHNOEVTAEEDGTORCEKSPCARVCGYGLMOYIKANSKEIGI 359
 QY 301 PYNVLTSDVGSCTLVCPLIHNOEVTAEEDGTORCEKSPCARVCGYGLMOYIKANSKEIGI 360
 DB 301 PYNVLTSDVGSCTLVCPLIHNOEVTAEEDGTORCEKSPCARVCGYGLMOYIKANSKEIGI 360
 QY 360 TELEFAGCKKIFGSLAFIPESFDDPASNNTAPLOPELOVETLEITGLYLSAMPDSL 419
 DB 360 TELEFAGCKKIFGSLAFIPESFDDPASNNTAPLOPELOVETLEITGLYLSAMPDSL 419
 QY 361 NVQEPDGGCKIFGSLAFIPESFDDPASNNTAPLOPELOVETLEITGLYLSAMPDSL 420
 DB 361 NVQEPDGGCKIFGSLAFIPESFDDPASNNTAPLOPELOVETLEITGLYLSAMPDSL 420
 QY 420 PDLISFONLQYIRGRILHNGAYSLLTQGLG:SWLGLRSLBELSGSLALIHNTHLCPVHT 479
 DB 420 PDLISFONLQYIRGRILHNGAYSLLTQGLG:SWLGLRSLBELSGSLALIHNTHLCPVHT 479
 QY 421 RDLISFONLQYIRGRILHNGAYSLLTQGLG:SWLGLRSLBELSGSLALIHNTHLCPVHT 480
 DB 421 RDLISFONLQYIRGRILHNGAYSLLTQGLG:SWLGLRSLBELSGSLALIHNTHLCPVHT 480
 QY 480 VPMQOLFRRPHQALLHTANRPEDE-CVGEGLACIQLCARHCHWCPGRTQCVNCSOPLRGQ 538
 DB 480 VPMQOLFRRPHQALLHTANRPEDE-CVGEGLACIQLCARHCHWCPGRTQCVNCSOPLRGQ 538
 QY 481 VPMQOLFRRPHQALLHTANRPEDE-CVGEGLACIQLCARHCHWCPGRTQCVNCSOPLRGQ 540
 DB 481 VPMQOLFRRPHQALLHTANRPEDE-CVGEGLACIQLCARHCHWCPGRTQCVNCSOPLRGQ 540
 QY 539 ECVEECRVLQGLPREYVNAHRCLEPCHECOPQNSVTCFGEADQCVACAHYKDPPECVA 598
 DB 539 ECVEECRVLQGLPREYVNAHRCLEPCHECOPQNSVTCFGEADQCVACAHYKDPPECVA 598
 QY 541 ECVEECRVLQGLPREYVNAHRCLEPCHECOPQNSVTCFGEADQCVACAHYKDPPECVA 600
 DB 541 ECVEECRVLQGLPREYVNAHRCLEPCHECOPQNSVTCFGEADQCVACAHYKDPPECVA 600
 QY 599 RCPBGVNPDLSTYMIKMFPEDEGACQPCPINCCHSCVDLDERGCPAORASPTVFIATV 650
 DB 599 RCPBGVNPDLSTYMIKMFPEDEGACQPCPINCCHSCVDLDERGCPAORASPTVFIATV 650
 QY 601 RCPBGVNPDLSTYMIKMFPEDEGACQPCPINCCHSCVDLDERGCPAORASPTVFIATV 660
 DB 601 RCPBGVNPDLSTYMIKMFPEDEGACQPCPINCCHSCVDLDERGCPAORASPTVFIATV 660
 QY 659 VGLILVAVLVGVFGLILKRRQKIRKTYMRLLOETVLEPLTPSGAMPQAOIRLIKET 718
 DB 659 VGLILVAVLVGVFGLILKRRQKIRKTYMRLLOETVLEPLTPSGAMPQAOIRLIKET 718
 QY 661 VGVLLFILVAVLVGVFGLILKRRQKIRKTYMRLLOETVLEPLTPSGAMPQAOIRLIKET 720
 DB 661 VGVLLFILVAVLVGVFGLILKRRQKIRKTYMRLLOETVLEPLTPSGAMPQAOIRLIKET 720
 QY 719 ELRKVKYLGSGAFNNFTVSWFLRV-----KVSASHLEVDRENTSPKANKELDEAVY 773
 DB 719 ELRKVKYLGSGAFNNFTVSWFLRV-----KVSASHLEVDRENTSPKANKELDEAVY 773
 QY 721 ELRKVKYLGSGAFNNFTVSWFLRV-----KVSASHLEVDRENTSPKANKELDEAVY 775
 DB 721 ELRKVKYLGSGAFNNFTVSWFLRV-----KVSASHLEVDRENTSPKANKELDEAVY 775
 QY 774 MAGVGSFYVSRLLIGICTSTVQVLTQMLPYGCLLDHVRNENRGLSGDLDLNMCMQIAKCM 833
 DB 774 MAGVGSFYVSRLLIGICTSTVQVLTQMLPYGCLLDHVRNENRGLSGDLDLNMCMQIAKCM 833

DB 776 MAGVGSFYVSRLLIGICTSTVQVLTQMLPYGCLLDHVRNENRGLSGDLDLNMCMQIAKCM 835
 QY 834 STLEDRLVHRRLAARNTVYKSPNVTITDPEGLALLDIDETRYHADGKVYIKMALES 893
 DB 834 STLEDRLVHRRLAARNTVYKSPNVTITDPEGLALLDIDETRYHADGKVYIKMALES 893
 QY 836 STLEDRLVHRRLAARNTVYKSPNVTITDPEGLALLDIDETRYHADGKVYIKMALES 895
 DB 836 STLEDRLVHRRLAARNTVYKSPNVTITDPEGLALLDIDETRYHADGKVYIKMALES 895
 QY 894 ILRRRTTHSDWASGVTVWELMTFGAKPYDGI PAREIPDLLEKSRLLPQPICTIDVYM 953
 DB 894 ILRRRTTHSDWASGVTVWELMTFGAKPYDGI PAREIPDLLEKSRLLPQPICTIDVYM 953
 QY 896 ILRRRTTHSDWASGVTVWELMTFGAKPYDGI PAREIPDLLEKSRLLPQPICTIDVYM 955
 DB 896 ILRRRTTHSDWASGVTVWELMTFGAKPYDGI PAREIPDLLEKSRLLPQPICTIDVYM 955
 QY 954 IMVKCMIDSECRPRELIVSEFSRMAEDPQRFVVIQNEEDCPASPLDSTFYRSLLEDD 1013
 DB 954 IMVKCMIDSECRPRELIVSEFSRMAEDPQRFVVIQNEEDCPASPLDSTFYRSLLEDD 1013
 QY 956 IMVKCMIDSECRPRELIVSEFSRMAEDPQRFVVIQNEEDCPASPLDSTFYRSLLEDD 1015
 DB 956 IMVKCMIDSECRPRELIVSEFSRMAEDPQRFVVIQNEEDCPASPLDSTFYRSLLEDD 1015
 QY 1014 MGLDVDAEEYLVPOGCFPCDPAPAGGVVHRRHSSSTRSGGGLTGLBESSEAPRS 1073
 DB 1014 MGLDVDAEEYLVPOGCFPCDPAPAGGVVHRRHSSSTRSGGGLTGLBESSEAPRS 1073
 QY 1074 PLAPSEAGSDVFDGDLGMAAGLQSLPTHDPSFLORYSEDPTVPLPSETDGYVAPLTC 1133
 DB 1074 PLAPSEAGSDVFDGDLGMAAGLQSLPTHDPSFLORYSEDPTVPLPSETDGYVAPLTC 1133
 QY 1134 SPQPEYVNOSEVQPPPLTSPGSLPPVRPAGATLTERATLSGKNGVYKGVAFGGAVEN 1193
 DB 1134 SPQPEYVNOSEVQPPPLTSPGSLPPVRPAGATLTERATLSGKNGVYKGVAFGGAVEN 1193
 QY 1194 PEYTLFOGGAAPQHPPPAFSPAFDNLVYDODPPEERGAPSTFTFGTPTAENPEYLGIDV 1253
 DB 1194 PEYTLFOGGAAPQHPPPAFSPAFDNLVYDODPPEERGAPSTFTFGTPTAENPEYLGIDV 1253
 QY 1254 PV 1255
 DB 1254 PV 1255
 QY 1256 PV 1257
 DB 1256 PV 1257

RESULT 3
 ERB2_MESAU
 ID ERB2_MESAU STANDARD; PRT; 1254 AA.
 AC 060553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
 GN ERB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 NC NCBI_TaxID=10036;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Neve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene";
 RL Gene 140:251-255(1994).
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE, GP130 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

QY 1076 APSEGASDVFDGDLGMAKGLQSLPTDHPSPLOQSESDPTVPLPSETDGYVAPLTCSP 1135
 DB 1076 APSEGASDVFEGLGMAKGLQSLPTDHPSPLOQSESDPTVPLPSETDGYVAPLTCSP 1135
 QY 1136 QPEYVNOQDVPDPSPSPREGPPLPAAPACATLERRATLSPGKGVKGVKVFAGGAVENPE 1195
 DB 1136 QPEYVNOQDVPDPSPSPREGPPLPAAPACATLERRATLSPGKGVKGVKVFAGGAVENPE 1195
 QY 1196 YLTPGGAAPQPPAPFSPAFNDLYYMDODPPERGAPESTFKGTAEENPEYLGIDVPV 1255
 DB 1196 YLTPGGAAPQPPAPFSPAFNDLYYMDODPPERGAPESTFKGTAEENPEYLGIDVPV 1255

RESULT 4
 EGFR_HUMAN STANDARD; PRT: 1210 AA.
 ID P05533; P05533; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00722;
 AC Q00688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (Receptor
 protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 expression of the amplified gene in A431 epidermoid carcinoma cells";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 MEDLINE=95382957; PubMed=7654366;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 epidermal growth factor receptor expression in human placenta";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 MEDLINE=97078686; PubMed=8938811;
 RA Reiter J.L., Mahle N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 factor receptor gene encodes a truncated form of the receptor";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 MEDLINE=97256547; PubMed=9103389;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 protein (TEGFR) in ovarian cancer";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX TISSUE=Placenta;
 MEDLINE=91100872; PubMed=1161793;
 RA Reiter J.L., Threadgill D.W., Riley G.D., Strunk K.E., Danielsen A.J.,
 Schehl Sinclair C., Parrall R.S., Green P.J., Yee D., Lampland A.L.,
 Balasubramanian S., Crosley T.D., Magnusen T.R., James C.D.,
 Mahle N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 mouse alternative EGFR transcripts encoding truncated receptor
 isoforms";
 RL Genomics 71:1-20(2001).

RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RX Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 Lampland A.L., Balasubramanian S., Crosley T.O., Magnusen T.R.,
 Mahle N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 extracellular domain of the receptor";
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 Evans R.M., Verma I.M., Gill G.N., Rosenfield M.G.;
 RT "Expression cloning of human EGFR receptor complementary DNA: gene
 amplification and three related messenger RNA products in A431
 cells";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 variety of RNAs overproduced in A431 carcinoma cells";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
 O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 receptor cDNA from human A431 carcinoma cells";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kilmington D., Ullrich A.,
 Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 identification of sequences regulating its transcription";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 transcript termination in the regulation of human epidermal growth
 factor receptor proto-oncogene RNA synthesis";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2391899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 epidermal growth factor receptor gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "Atp-stimulated interaction between epidermal growth factor receptor
 and supercoiled DNA";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Domagalian M., Honegger A.M.,
 Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)

RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor."
 RL J. Biol. Chem. 264:10667-10671 (1989).
 RN [16]
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; Pubmed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts."
 RL Growth Factors 13:121-132 (1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; Pubmed=10721668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor."
 RL J. Biochem. 127:65-72 (2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; Pubmed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Iax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor."
 RL J. Biol. Chem. 273:11150-11157 (1998).
 RN [19]
 RP REVIEW.
 RX MEDLINE=87297456; Pubmed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens."
 RL Annu. Rev. Biochem. 56:881-914 (1987).
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X00588; CA28240.1; -
 DR EMBL: U95089; AAB53063.1; -
 DR EMBL: U48722; AAC50802.1; -
 DR EMBL: U48723; AAC50804.1; -
 DR EMBL: U48724; AAC50796.1; -
 DR EMBL: U48725; AAC50797.1; -
 DR EMBL: U48726; AAC50798.1; -

QY 11 LLLALLPGAA--STQVCTGTDMLRLPASPEHLDMRLHLYGCGVQVGNLELYLPTN 68
 DB 14 LLAALCPASRALEEKVKCGSTSNKLTQGTGFEDHFLSLORMENNECVLTGNLEITYQGN 73
 QY 69 ASLSPFDIOIEQGVYVLAHNOVQVRLRLIVNGTQGFEDNYVLAVIDNDPLNTTP 128
 DB 74 YDLSFLKTIQEVAGVYVLAHNTVERIPLENQIIRGNMYENSVLAVALSNYD----- 126
 QY 129 VTGASPGGLREQLSLTEILKGVLIQRNPQCYQDTIILMDIFHNKQLALTIDTNR 188
 DB 127 ---AKTGKLEPMNLQELHGAIRFSNRPALCNVBSIQMRDIISSDLSMSMSDFQNH 183
 QY 189 SPACPCSPMCWCKSGCKWGESSEDQSLTRTVACGCA-RCGKPLPTDCHECAAGCTCP 247
 DB 184 LSCCKCPSCPCNGSCWCGENCCOKLTIKCAQCCSRCKRCKSPSDCHNCAAGCTGP 243
 QY 248 KASDGLACLFHNSGICELHCPALVTYNTDFESMPNDEGRATFASACTAPRYVLSND 307
 DB 244 RESDGLVCKRFDEATCKDTCPLMLYPTTYQMDVNEBKISFQATCYKCPRYVYTD 303
 QY 308 VGSCTLVCPLANQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGITTELE-FAG 366
 DB 304 HSGCYRACGADSEYEM-EEDGVYKCKCKCGPCGKVCNGI:GIGEFK-DSLSINATNFKHFN 361
 QY 367 CKKIGSLAFIPESFDGPASTAPLQEQLOVFTLEITGYLIYSAPDSLPULSVFQ 426
 DB 362 CTISGDLHLIPVAERGDSFHTPPLDQELDLITVRELITGFLIQAMPENRTDLHAFE 421
 QY 427 NLQVIRGRILHNGAVSLTQGLIGISWLGIRSLRELSGALIHNTHTCFVTVPMDOLE 486
 DB 422 NLEIRGRTKHQGQSLAVSNLITSLGRSLKESDSDVILSGKNLCYANTIMKKILF 481
 QY 487 RNPHQALLHTANRPDECEVGEGLACHQLCARHGCKGPGPTQCVNCSQFLRQGECEYECRV 546
 DB 482 GTSGCKTKIISNRGNSCKATQGVCHALCSBEGCKPBPBCVCSGRNRSRGECDVCKNL 541
 QY 547 LQGLPREYVNAHCLPCGPECOPNGSTCGPFEADQCVACHYNDPFCVARGCSGVKP 606
 DB 542 LEGERREVENSECTQCPBELPQAMNNTITGREGDNCIQCAHYIDGPHCVAGVWG 601
 QY 607 DLSYPIWKFPDEBACQPCPINTCHSCVDLDKCPAEOBASPLTISVSAVG--ILL 663
 DB 602 ENNTL-VMKYADAGVCHLCHPNCTYGTGPGLECCPENGKIP--SLATGMVGAALLLL 658
 QY 664 VVVLGVVGLIKRQOKIRKTYMRLLQETELVBPILPSCAMPNQAOMRLIKETELRV 723
 DB 659 VVALGIG--LFMRRHIVRKRTLRLQEBRELVEPLTPSGEAPPAQALLRIKETEFPKI 715
 QY 724 KYLSGAGFNNFTVSFMLRVPKVSASHLEV---LRNTSPKANKETLDEAYVWAGVS 779
 DB 726 KYLSGAGF---TYRKGLMIPBEGKVKIPVAKIKELREATSPKANKETLDEAYVWASVDN 771
 QY 780 PYVSRLLGILSTVQVLTQMLPYGCLLDHVRNKGRLGSDLLNMCQIAKAGMSYLDV 839
 DB 772 PHVCLLGICITSTVQLTQMLPGCLLDVYREHNDNGISGYLLNMVQIAKAGMYVLEDR 831
 QY 840 RLVRHDLAARVNVKSPHNVKITDPCGLARLLDIETEVHAGGVKIPFMWLESTLRRF 899
 DB 832 RLVRHDLAARVNVKTPHVKITDPCGLAKLGAEEKEVHAGGVKIPKMWLESTLHRTY 891
 QY 900 THQSDVMSYGVTVMLMTFGAKPYDGI:PAEIPDLLEKGEELPQPICTIDVMIWIMYK 959
 DB 892 THQSDVMSYGVTVMLMTFGSKPYDGI:PAISISSILEKGEELPQPICTIDVMIWIMYK 951
 QY 960 MIDSCRRFRELIVSEFRRARNDQRFVIO-NEDLQFASLDSFTFRSLLEDMDGDLV 1018
 DB 952 MIDASRRFRELIIIESKARNDQRYLVIVIGDERMHLPSPTDSNFYALMDEEDMDVDV 1011
 QY 1019 DAEELYVQGGFCFPCDPAPAGAGVWHHRHRSSTRSGGDLTLGLESEEBEAPSPAPLS 1078
 DB 1012 DADETLIPQGGF-----SSPETSRTPLLS 1037

Query March 44.1k; Score 3007; DB 1; Length 1210;
 Best Local Similarity 48.6k; Pred. No. 9e-154;
 Matches 621; Conservative 169; Mismatches 367; Indels 120; Gaps 27;

FT DISULFID 620 628 BY SIMILARITY.
 FT DISULFID 624 636 BY SIMILARITY.
 FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL).
 FT CONFLICT 19 19 C -> S (IN REF. 2).
 FT CONFLICT 539 539 C -> W (IN REF. 5).
 FT CONFLICT 991 991 L -> F (IN REF. 4).
 FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
 SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD2F5 CRC64;

11 LLLALLPPGAA--STOVCTGTDMKRLPASPEHLMRLHLYOGCQVQGNLELYLPTN 68
 14 LTRALCAAGALEKVKVCCGTSNRLLTGLTFEDHFLSLQRYMNNCEVLNLLETIVORN 73
 69 ASLSFLDIDIOVQVYVIANOVROVLOSLRITRGOLFEDYALALVINDGPNLNTTP 128
 74 YDSFLTKIDVAGVYVIALNTVERIPLEMLQIIRGALYENTYALALISN----- 124
 129 VTGASPGRLRELQSLTEILKGGVLIQRNPOLCYOQTILMKDI---FRRNQLALTLI 184
 125 -YGNRTGLRELPRNQLGELIGAVRFSNNPILCMNMTIQRDIQVQVFNSSMNL--- 180
 185 DTNRBRACHPCSPCKSRCKSSSEEDCQSLTRVACGGA-RCKEPLPTDCCHEQCAAG 243
 181 -QSHSSCPKCDPCSPGSCWGGEENCQKLTIKICQOCSHRCKRSPDCCCHNOCAAG 239
 244 CTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPESMNPREGRYFFGASCYACAPYV 303
 240 CTGPRSDCLVCCQFQDEATCKTCPLMLNPTTYMDVNPBCKSKSPGATCYKCKPRNY 299
 304 LSTDVSGCTLVCPLNQVTAEDGTQRCCKSKPCACVCGLMQYIKANSKFIITEL 363
 300 VVTDHSGCVACGDEYEV-EDGIRKCKKDDGFCRVACNGIGIEFR-DTLSINATNIK 357
 364 -FACCKKIFGSLAPESFDGDPASNTAPLOPELOVFELEITGLYISAPDSAPDL 422
 358 HFKCTAISGDHLPLVAFKSDSTRTPLDPRIELTKVETIGLLIQAPMDWTDL 417
 423 SVFONLOVIRGRILHNGAYSLTLOGLISWGLSRLELSGGLIHNHNLGFVHTPW 482
 418 HAFENLEIRGRITQHGQFSLAVGLNITSLGLRSLKEISDGVYIISGNRLCYANTINW 477
 483 DQLFNRPHOALHNANPDECEVGEGLAQHQLCARGCKNPGPPOCCNCGQFIRGQECVE 542
 478 KKLFGTNPQTKIMNNPAEKDCAVNHVNCPLSSSECCWGPEDDCSCONVSRGRCVE 537
 543 ECRVLQGLPREYVNAARCLPCHPECCQFQNGSVTCFGEADQCAVCAHYKPPFCVAPCP 602
 538 KCNILBEPREFEVNSECICHPRECLPQAMNITLTGSPNQCQAHYIDGPHCVKTCRA 597
 603 GVAPDLSTYPIWKPPDEGACQPCPINCHTSCYDLDKGPAPORASPLTISAVVIGIL 662
 598 GIMGENTNL-VMKYADANNVCHLHANCATGACGPGIGQCEVWPSGPKIPSIATIGVIGL 656
 663 LVVVLGVVFGI-LIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKELELR 721

DB 657 LPIVAVLQIGLFFRRRHIVKRLRLDERELVEPLTSPGAPNQAHLKLETERK 715
 QY 722 KKVYVSGAFGFNFNTVFWLVPKVSASHLEV---LRENTSPANKREILDEAVYMAV 777
 DB 716 KKVYVSGAFG---TVYKGLMTPEGEKPKIPVALKELREASPANKREILDEAVYMAV 771
 QY 778 GSPYVSRLLIGLITSTVOLVTQMPYGCLLDHNRENRRLSGQDLLNMCQIAKMSYLE 837
 DB 772 DHPHVCRLRLLIGLITSTVOLITQMPYGCLLDHNRENRRLSGQDLLNMCQIAKMSYLE 831
 QY 838 DRLVLRDLAANVNVKSPNHYKIDPGIARLDIDENRYADGGKVPKMALESILHR 897
 DB 832 DRLVLRDLAANVNVKSPNHYKIDPGIARLDIDENRYADGGKVPKMALESILHR 891
 QY 898 RFTTHSDVWSYGVTVWELMTFGAKPEYDIPAREIPDLLEKGERLPQPPICITDYVYIMWK 957
 DB 892 ITHQSDVWSYGVTVWELMTFGAKPEYDIPAREIPDLLEKGERLPQPPICITDYVYIMWK 951
 QY 958 CMWITSECCPRRELTVSEPSRMAPDPORFVYQ-NEDIGAPASPLDSTPYRSLLEDMDMD 1016
 DB 952 CMWITSECCPRRELTVSEPSRMAPDPORFVYQ-NEDIGAPASPLDSTPYRSLLEDMDMD 1011
 QY 1017 LVDAEEYLVPOQGFPCPDPAFGAGVHRRHSSSTRSGGDLTLGLEPSEEARPSPLA 1076
 DB 1012 VVDADEYLVPOQGFPCPDPAFGAGVHRRHSSSTRSGGDLTLGLEPSEEARPSPLA 1071
 QY 1077 PSEGAGSDVFDGDLGMAAGLQSLPTHPSPLOQYSEDPTVPLPSET-DGYVAPLTC 1134
 DB 1038 SLSLATSNT---NSTVACINRGSCRVKEDAPLQYSSDPTGAVTEDNDIAFL----- 1087
 QY 1135 PQPEYVNPDPVPPQSPREPGPLPAPRPAATLEAKTLSPKNGVNVDFVFGAVENP 1194
 DB 1088 PVPEYVNPQ-SVPRKPRAGSVQNPVHNQPLHP-----APGRDLHYQN-DHSAVAGVP 1136
 QY 1195 EYL-TPQGAAPQPPPPAPFSPAFNLYWMDQ-----DP-----PERGAPSPTF 1237
 DB 1137 EYLNIAQ-----PCLSSGFNSPALMIGKSHQMSLDNDVDVQDQFFPRETPNGIF 1187
 QY 1238 KGTPTAENPEYGLDVP 1254
 DB 1188 KG-PTAENPEYGLDVP 1203

RESULT 6
 ERB4_HUMAN
 ID ERB4_HUMAN STANDARD; PRT; 1308 AA.
 AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=breast carcinoma;
 RX MEDLINE=93189574; PubMed=8383326;
 RA Plozman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
 Foy L., Neubauer M.G., Shoyab M.,
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
 RT epidermal growth factor receptor family".
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).
 RC TISSUE=fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Cortes G., Paul S., Choi C.J., Rio C., Plozman G.D.,
 RA Klagesbrun M.;

"A novel juxtamembrane domain isoform of HER4/ErB4. Isoform-specific tissue distribution and differential processing in response to phorbol ester".
 J. Biol. Chem. 272:26761-26768(1997).
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: JM-A (SHOWN HERE) AND JM-B; ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART, KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM, PLUTINARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS, LUNG, SALIVARY GLAND, AND PANCREAS.
 CC -1- PPM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: L07868; AAB59446.1; -
 DR HSSP: P1362; IFGK.
 DR Gene: HGNC:3432; ERBB4.
 DR MIM: 600543; -
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; YLP_Motif.
 DR Pfam: PF00069; Pkinase_1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR Pfam: PF02757; YLP; 2.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SMO0261; FU; 4.
 DR SMART: SMO0219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KM Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 1308
 FT DOMAIN 26 651
 FT TRANSMEM 652 675
 FT DOMAIN 676 1308
 FT DOMAIN 186 334
 FT DOMAIN 496 633
 FT DOMAIN 718 985
 FT NP_BIND 724 732
 FT BINDING 751 751
 FT ACT_SITE 843 843
 FT DISULFID 189 197
 FT DISULFID 193 205
 FT DISULFID 213 221
 FT DISULFID 217 229
 FT DISULFID 230 238

FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 259 BY SIMILARITY.
 FT DISULFID 262 289 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1168 1168 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).
 FT VARSPLIC 626 648 NGPTSHDCIYFWTGHSTLPQNA -> IGSSIEDICGLMD (IN ISOFORM JM-B).
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;
 Query Match 41.9%; Score 2852.5; DB 1; Length 1308;
 Best Local Similarity 44.1%; Pred. No. 18e-145;
 Matches 596; Conservative 187; Mismatches 385; Indels 183; Gaps 31;
 QY 9 WGLLLALLPPGAA---STOVCTGDMKRLPASEPETHLMDLRHLYOGGVYQGNILETY 64
 DB 8 WTVVSLVLAAGTVQSDSQSCAGTENKLSLSDEQFALRKYNEVEVMGNLEITS 67
 QY 65 LPTNASLFLQDIOEVGVYLAHQVQVPLQRLRYRGOLFEDNYALALVDNDPLN 124
 DB 68 IEHNDLSFLRSVRETVGLVALNQFRLPLENLRITRGTYLVEDRVALALFLYRKDG 127
 QY 125 NTPPYTGAAPGGLREIQLRLTEILKGVLIQGNPDLQCYODTILMKDIFHKNOQLATLI 184
 DB 128 NF-----GLQELGLKNLLEILNGGVYVQNNFLCYADTIHQDIYRNWPENLTLV 178
 QY 185 DTRNSPACHPCSPMKGSGFCWGESSEDCSLTRTYCAGGC-ARCKGPLETDCHEQCAG 243
 DB 179 STNGSSGCCRCHKSGCTG-RCWGPTENHCOTLTRTYCAECDCRCGPFYVSDCHRECAAG 237
 QY 244 CCGPKSDDCLAHFHSGIGCELHCPALVTYTDTPESMPNDEGRYTGASCVTACPFY 303
 DB 238 CGSPKDTDFACMKNFSDSACVTCQPPOTFYVNPPTTQLEHNNAKTYTGAFVCKACPHNF 297
 QY 304 LSTDVSGCTLVCPLEHNOEYTABDGTORCEKSKPCARVYCYGIMQYIRANSKFTIGTELE 363
 DB 298 V-VDSSSCYRACPFSSIMEY-ENGIGKMCPCDTCIPKACDGIIGTSLMSAQVDSNDIK 355
 QY 364 FAGCKKIFSGSLAFESFGDPDPASTAPLPOLQVFELEITQVLYISAMPDLSPLS 423
 DB 356 FINTCKINNLFLVTGHHGDPYNAIEADPEKLVFRTVREITFPLINQSPFPMWTPDS 415
 QY 424 VFQNTQVIRGRILHNGAVSLTQGLGISNLTGLRSLEISGIALITHNTHLCFYATVMD 483
 DB 416 VPSNLVTIGRVLVYSGSLILKQGGITSLQGSLEISAGNIYITDMSNLCYHTIMWT 475
 QY 484 QLFRRPHQALLHTANRPDECVGEGLACQLCARHGCMGPGPTQCVCNCSQPIRGECEVBE 543


```

Db      476 TLFSTINRIIVRDNRAKNAETAENAVCNHLCSSDGCWGPBGDOCLSCRRFRGRICIES 535
Qy      544 CAVLGQLPREYVNAHCHLPCHEPCOP-CNGSVTCGFEPAADQCVACAHYKDPFPCARCPSS 602
Db      536 CNLYOGEPFEFENGSIQVCECPQCEKMEGGLTCHGPPGPDNCTKSHKRDGNCVCEKPPD 595
Qy      603 GVKPDLISYMPWIKPFDEGACQPCPINCTHSCVDDDKGC-----PAEQASPL 651
Db      596 GUGANSF--IFKYADPDRECHCHPHNCTGCGNGPFSHDCIYYPWTGSHLFGHAR-TPL 652
Qy      652 TSVASVY-GILLVVLGVVPGILLKRRQOKIRKTMRLLOETLVEPLTPSGAMPQA 710
Db      653 --LAAGVIGGLFVLIVGLTFAVYVRKSIK-KKALRFL-ETLVEPLTPSGAMPQA 708
Qy      711 QMRILKETELRKVKVYLGSGAFGPNFTVSFWLRVP-----KVSASHLEVRENTSPKXNK 765
Db      709 QRIIKETELKRVKVLGSGAFG---TYKKG:WVEGETVKIPVA-IKILNETTPKXNV 763
Qy      766 ELIDENYVWAGVSPYVSRLLGICLTSTVQVLTQMPYGCILDHREKRGRLGSDLLNW 825
Db      764 EFMDEALIVASWDPHLVRLGLVCLSPITQVLTQMPHGLLVEYHEKDNIGSQLLLNW 823
Qy      826 CMQIAKMSYLEDELVLRDLAARVLYKSPNHVKITDGLARLLDIDETEHADGKVP 885
Db      824 CVQIAKGMVYLERLRVLDLAAARVLYKSPNHVKITDGLARLLDIDETEHADGKVP 883
Qy      886 IKWMALESLRRRFTHOSVWGYVTWELMTFGAKPYDGPAREIPDLLEKGERLPQP 945
Db      884 IKWMALECHYHRKFTHOSVWGYVTWELMTFGKPYDGPAREIPDLLEKGERLPQP 943
Qy      946 ICTIDVYIMVCKWIDSECRPRELWSEFSRMDQRPFVIGNEO-LGASPLDSTF 1004
Db      944 ICTIDVYIMVCKWIDSECRPRELWSEFSRMDQRPFVIGNEO-LGASPLDSTF 1003
Qy      1005 YKSLLEDMDGLVDAEYLVQOGFCPPDPAQAGWVHHRSSSTRSGGDIYLLGLE 1064
Db      1004 FQNLDEBDLMDMAEYLVQ-CAFNIPPP-----IYTSRARIDSKRS-----EIGH 1051
Qy      1065 PSEEDAPRS-----PLAP-SEAGSDVFDGDLGMMAA 1095
Db      1052 PPVATPWSGNOFYRRDGGFAAEQGVSPYRAPSTITFEPAPAQATKTEIFDSCCNLT 1111
Qy      1096 KGLQSLPTHPDPSLPLOYSDEPTVPLPS-----ETDGVVAPLTGSPQEVYNQPDVAPQ 1148
Db      1112 KRPVAPVHGEDSSSTRYSADPTVFAPERSPRGELEDEGMTPMRKQEVLYNPVE---- 1167
Qy      1149 PPSPREGLPAPRPPAGATLERAKTILSPGKNGVYKDVFAFGAVENPEYLTPOGGAAPQH 1208
Db      1168 -----ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG----- 1194
Qy      1209 PPPA-----FSPAFDLVYMDQDPPRGA-- 1232
Db      1195 PPKADEVYNEPLVYNTFANTLIGKAEYLNKNILNLPKAKKAFDNPYWNHSLPPRSTLQ 1254
Qy      1233 PPSTFKGTP-----AENPEYL 1249
Db      1255 HEDVLOEYSTKYFYKNGRIRPIVAENPEYL 1285

```

```

RP      SEQUENCE FROM N.A.
RC      TISSUE=Heart;
RX      MEDLINE=98221155; PubMed=9553078;
RA      Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA      Marchionni M.A., Kelly R.A.,
RT      "Neuregulin promotes survival and growth of cardiac myocytes.
RT      Persistence of Erb2 and Erb4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      [2]
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=91222560; PubMed=2025425;
RA      Lai C., Lemke G.,
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system."
RL      Neuron 6:691-704(1991).
RN      [3]
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.,
RT      "Expression of neuregulin and their putative receptors, Erb2 and
RT      Erb3, is induced during Wallerian degeneration."
RL      J. Neurosci. 17:1642-1659(1997).
CC      -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NNAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERB
CC      RECEPTORS (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, ADI
CC      HEART.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.sib.ch).
CC      -----
DR      EMBL; AF041838; AAD08899.1; -
DR      EMBL; U52531; AAC53051.1; -
DR      HSSP; P11362; IFGK.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_kinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_kinase.
DR      InterPro; IPR004019; VLP_motif.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF00069; pkinase; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      Pfam; PF02757; VLP_2.
DR      PRINTS; PR00109; TYRKINASE.
DR      PRODOM; PD000001; Euk_kinase; 1.
DR      SMART; SM00261; Fv; 4.
DR      SMART; SM00219; TYRK; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT      SIGNAL
FT      1
FT      25
FT      POTENTIAL.

```

FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 652 675 POTENTIAL.
 FT DOMAIN 676 1308 CYTOSOLASTIC (POTENTIAL).
 FT DOMAIN 186 334 CYS-RICH.
 FT DOMAIN 496 633 CYS-RICH.
 FT DOMAIN 718 985 PROTEIN KINASE.
 FT NE BIND 724 732 ATP (BY SIMILARITY).
 FT BINDING 751 751 ATP (BY SIMILARITY).
 FT ACT_SITE 843 843 BY SIMILARITY.
 FT DISULFID 189 197 BY SIMILARITY.
 FT DISULFID 193 205 BY SIMILARITY.
 FT DISULFID 213 221 BY SIMILARITY.
 FT DISULFID 217 229 BY SIMILARITY.
 FT DISULFID 230 238 BY SIMILARITY.
 FT DISULFID 234 246 BY SIMILARITY.
 FT DISULFID 249 258 BY SIMILARITY.
 FT DISULFID 262 269 BY SIMILARITY.
 FT DISULFID 293 304 BY SIMILARITY.
 FT DISULFID 308 323 BY SIMILARITY.
 FT DISULFID 326 330 BY SIMILARITY.
 FT DISULFID 503 512 BY SIMILARITY.
 FT DISULFID 507 520 BY SIMILARITY.
 FT DISULFID 523 532 BY SIMILARITY.
 FT DISULFID 536 552 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 559 577 BY SIMILARITY.
 FT DISULFID 580 589 BY SIMILARITY.
 FT DISULFID 593 614 BY SIMILARITY.
 FT DISULFID 617 625 BY SIMILARITY.
 FT DISULFID 621 633 BY SIMILARITY.
 FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 138 138 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 1062 1062 S -> N (IN REF. 3).
 FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
 SQ SEQUENCE 1308 AA; 146957 MM; D944B096A08B41 CRC64;
 Query Match 41.64; Score 2833; DB 1; Length 1308;
 Best Local Similarity 43.94; Pred. No. 2e-144;
 Matches 594; Conservative 194; Mismatches 394; Indels 170; Gaps 31;

QY 297 TACPYNLTSTVSGCTLVCPPLHNOEYTAEDGTORCEKSKRCARVCGYLGMOYIKANSKF 356
 DB 291 KCCCHENFV-VDSSCCVACAPSSKMEV-ENQIKCKKCTDTCRACDGIIGSLMSQTV 348
 QY 357 IGTELEFAGCKKIFGSLAPLPEFDDPASNTAPLOPELOVETLEITGYIYISAMP 416
 DB 349 DSSNDKRNCTKINGNIFLVTGIGHGPYNAIDIPBEKLVNRTYREITGFLINQTPW 408
 QY 417 DSLPELVSFQNLQVIRGILHNGAVSLTLQGLGSIWGLRSLRGLSGLALIHNTHTCF 476
 DB 409 PMNDFSVFSLVLTGGKVLVSGSLTLKQGLITSLOFSLKEISAGNIYITNSNLCY 468
 QY 477 VHTVPMQDLFRNPQALHTANREDECVSGEGLAGHOLCARGHMGCPPTCVANCSQFLR 536
 DB 469 YHTINMTLFTSTVQRIYIRDNRAENCTAGAWCNHLCSDGCMGPGDCLSCRPSR 528
 QY 537 GQECYEBEYVLOGLPREVYVARHCLPCHPEQCP-QNSVTCFGEADQVCAYKXDPF 595
 DB 529 GKICIESCNLYGDEFERENGISICVECDSDQCKXKEDGLLTGHGPPNCTKSHFKQGN 588
 QY 596 CVARCPGVKPDLSYMPIMKPEDEGACQCPICNCTSHCVLDLQKGC-----PA 644
 DB 589 CVCKCPDVLQGANF--IFKYADQDRCHPCHPCTGCGNPTSHDCLYPTWGTSTLPQ 646
 QY 645 EQRASPLTSIYSAV-GLLVVGLGVVFGILIKRQOKIRKRYMRLLQETELVEPLTS 703
 DB 647 HAR-TPL--IAAGYIGGLFIVIMALTFAVYVRKRSIK-KRALRRFL-ETELVEPLTS 701
 QY 704 GAMPQAOAMRLKTELRLKRYVVGSGAFGNFVSWLVRP-----KVASHLEVREN 758
 DB 702 GTAPQAOALRLKTELRLKRYVVGSGAFG---TVYGIWVPESETYKIPVA-IKILMET 756
 QY 759 TSPKANKEILDEAYVMAGVGSPPYVSRLLGICLTSTVQVLQMLPEYGCULDHVRNRRGL 818
 DB 757 TGPANVEMFDEALIMASVDRPHLVRLGVCLEPTIQVLQVLMHGGCLEVHEHKNIG 816
 QY 819 SQDLINMCQIAKMSYLEDVLRHDLAANVLYKSPNNHYKITDFGLARLLDDEFEYH 878
 DB 817 SQLLIMCVQIAKMMYLEERLVHRDLAANVLYKSPNNHYKITDFGLARLLDDEFEYH 876
 QY 879 ADGGVPIKMALESILRRFTHSDVSVYVWELMTFGAKPYDGIPIAREIPDLLEKG 938
 DB 877 ADGGMPFKMALEIEIHRKFTHSDVSVYVWELMTFGAKPYDGIPIAREIPDLLEKG 936
 QY 939 ERLPPICTIDVTYIMVYKMMIDSECPRELVSEFSRYARDPQRFVYIQNED-LGPA 997
 DB 937 ERLPPICTIDVTYIMVYKMMIDSECPRELVSEFSRYARDPQRFVYIQNED-LGPA 996
 QY 998 SPLDSTFYRSLLEDDMDGLVDADAEYLVPOQGFCCPDP----- 1035
 DB 997 SPNDKFFQNLDEDELDMDADAEYLV-CAFMIPPEIYTSRTRIDSNRSEIGHSPPPA 1055
 QY 1036 -APGAGVHHRHRSSTRSQGDULTGLEPSEEARPSPLAPSEGAGSDVFDGLMGCA 1094
 DB 1056 YTPMSGQFVYQDGGFATQGC--MPMYTATSTIPAPVA--QGATAEMFDDSCNGT 1110
 QY 1095 AKGLQSLPHTPSPLOKRSDEPTVPLRS-----ELDGYAPLTCSPQRYVAVQDVRP 1147
 DB 1111 LRKPVPHVQDSSSTORSADPTVFABERNRAELDEGYTTPMHDREKQYLPVVE-- 1167
 QY 1148 QPSPREGEPLPAAPAGATLERAKTLSPGKAGVAVDFAGAVENREYLLPQGAAPQ 1207
 DB 1168 -----ENPFVSRK-----KNGDLO-----ALDNEIEYHASSG----- 1194
 QY 1208 HPPPA-----ESPADFNIYWDQDPPEGA- 1232
 DB 1195 -PPRADEYVNEPLYLNTFTNALGNAEYMKSLLSVEBKAKKADNDNDYMHSLPSTLT 1253
 QY 1233 -PPSTFKGTP-----AENPEYL 1249
 DB 1254 QHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 8
XMRK_XIPMA STANDARD: PRT: 1167 AA.
ID XMRK_XIPMA
AC P13388;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
GN XMRK OR TY.
OS Xiphophorus maculatus (Southern platyfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphi; Acanthopterygii; Percomorpha; Atherinomorpha;
CC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OK NCBI_TaxID=8083;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90015140; PubMed=2797166;
RA Witbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
RA Telling A., Robertson S.M., Scharlt M.;
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
inducing tu locus in Xiphophorus.";
RL Nature 341:415-421 (1989).
RN [2]
RP REVISION TO 515.
RA Scharlt M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X16891; CAA34770.2; -.
DR PIR; S06142; S06142.
DR HSP; P11362; IFGK.
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR000719; Euk_Dkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_Chr_Pkinase.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 5.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KM Tyrosine-protein Kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 25
FT CHAIN 26 1167
FT DOMAIN 26 642
FT DOMAIN 643 665
FT DOMAIN 666 1167
FT DOMAIN 710 977
FT NP_BIND 716 724
FT BINDING 743 743
FT ACT_SITE 835 835
BY SIMILARITY.

FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 228 BY SIMILARITY.
FT DISULFID 224 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 311 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 333 337 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 508 521 BY SIMILARITY.
FT DISULFID 524 533 BY SIMILARITY.
FT DISULFID 537 553 BY SIMILARITY.
FT DISULFID 556 569 BY SIMILARITY.
FT DISULFID 560 577 BY SIMILARITY.
FT DISULFID 593 615 BY SIMILARITY.
FT DISULFID 618 626 BY SIMILARITY.
FT DISULFID 622 634 BY SIMILARITY.
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1167 AA; 129934 MW; 479380749DCID55A CRC64;
Query Match 37.8%; Score 2575.5; DB 1; Length 1167;
Best Local Similarity 44.0%; Pred. No. 1,1e-130;
Matches 561; Conservative 168; Mismatches 392; Indels 155; Gaps 33;
QY 4 AACRCGLLALLPPEAST---QYCTGDMKRLRPAPEHLMRLHYGGCYVQGN 59
DB 8 ALLIQ--LLVLSISRCSTDPDRKVCQSTSNQMTN--LDNHYLMKKMYSGCCNVLEN 62
QY 60 LELTYLPINASLFLDIDIEVQGYVLIANQVQVPLQRLRVGTOLPFDNYVALVDN 119
DB 63 LEITYQENQDLSFLQSIQEVGYVLIANNEVSTIPLVRLIRGQNLVEGFLLVMEN 122
QY 120 GDLNNTPTVYASRGLRGLRELQRLSTETLTKGVLIQRNPQLCYQDTILMKDIFRNNOL 179
DB 123 YQR-NPSSP--DVYQGLKQLQSLNLTETLSGVKVSHPLLCNVETIMMDIVXTSNP 179
QY 180 ALLIDITNRSRACHPSPKSGSRGSESEDCQSLTRTVCAAGC-ARCKGLPTDCCH 238
DB 180 TNNLIPHAFERQCKDHCQVAGSCMAPRGHCQKFTLLCAEQCNRRRGKPIDCCNE 239
QY 239 QCAAGCTGKASDCLACHFNHSGICELCPALVYNTDTFESMPNPEGRTYFGASCYTA 298
DB 240 HCAGGCTGPRATDCCACRDYFNDGCTCKDPPKXIYDIYSHQVVDNPNKXIFGAACVKE 299
QY 299 CFYNYLSTVGSCTLVCP-AMNQEVTAEDTQCEKSCRKCAVCAGLGMQYIKANSKFTG 358
DB 300 CSNRYVTE-GACVRSACSGMLEVD-ENGKSKCKCDGVCPVCGIGGSI-SVTIAVN 356
QY 359 ITEL-EPAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQYFETLEETITGLYISAMP 417
DB 357 STNINSFSNCTKINDIILNRSFEGDPHYKIGTMDPELMLNLTIVKEITGLVIMWPE 416
QY 418 SLPLDSYQNLQVITGRILHNAYS-LTLQIGISWLGIRSRBELSGALIHNTLCLF 476
DB 417 NNTSLSVFQNLLEIIGRTTFSGFSFYVQVYHMLGIRSRKEVSAGVAILKNTLQRY 476
QY 477 VHTVPMDQLFRNPQALHTANRPDECEVGEGLACHQLCARGHCVPQTCVNSQGLR 536
DB 477 ANTIMRRLRFREDSISVDART-----ENQTCNNBCSBEKCGPPTKVCGLAHDR 529
QY 537 GGECEVBEKRVLQGLFREYVNAHCLPCHPBCQPNQNGSVTCFPEADQCVACAHYKDPFPC 596

```

Db 530 GGRCVASCNLLQGEPRBAQVGRVQCHQCECLVQDLSLTCVGPFPANCSKSAHPQDQJC 589
Qy 597 VARGSSGVPRDLSTYPMKFPDEBACOPCPINCHSCVDDDKCCPRAEPASPLTIVS 656
Db 590 IPRCPGILGDDTL-IMKADKWCQCPCHONCQSGPGLSGCRD-IVSHSLAVG 647
Qy 657 AVVGLLVVGLVGFILKRQCKIRKYMRLQETELVEPLTPSAMPPOAMRLK 716
Db 648 LVSGLLITVALLVLLVLRRLRRK-RKRTICLLQKELEVEPLTPSOANQAFRLTK 706
Qy 717 ETELKRVKVLGSGAGCFNNFTV-SFW-----LAVPKVSAHLEVLRENTSPKANKEL 768
Db 707 ETEFKKRVLGSGAGG---TYXKGLMPDGENIRIPVA---IVLEAETSPKVNQEV 758
Qy 769 DEAYVAGVSPYVRLGICLTSTVQVLTQMLPGCLLDHYRENRGLSGODLLNMCQ 828
Db 759 DEAYVMSVDHVCRLGICLTSAVOLVLTQMLPGCLLDYRQGEICQOMLNCVQ 818
Qy 829 IAKGMSYLEDVLRHDLAARNVLYKSPNHVKTDFGLARLLDIDETBYHADGKVPK 888
Db 819 IAKGMVYLEERLHRLDLAARNVLLKNPNHVKITDFGLSKLTADKEKYOAGDGKVPK 878
Qy 889 MALESILRRRTFQSDWSYGVTLWELTFGAKPYDGI-PAREIDPLEKGRLOPPIC 948
Db 879 MALESILQWTTTHQSDWSYGVTLWELTFGSKPYDGI-PAKEIASVLENGERLPPIC 938
Qy 949 IDVYIMVYKQWIDECRPRFRELVEFSRMAIDQREVIQNDLGPASPLDSTFYSL 1008
Db 939 IEVYMIILKQWIDSSRPREFELVGFSGMARDSRYLVIGQ--NLPSLSDRRLSRL 995
Qy 1009 LEDDMGDLVDAEELVLPQGGFCPCDPAPAGAGVHHHRSSSTSGGDLTLGLEPSE 1068
Db 996 LSSD--DVADDEVLLPYKRI-----NFGGS----- 1020
Qy 1069 EAPRSPAPSEAGSDVDFDGLMGAKGLOSLPHTDPSPLQRYSEDTV-PLPSETGY 1127
Db 1021 -EPCIPPTGH-----P-RENSITLRNISDTQNALEKDLGH 1056
Qy 1128 VAPLTCSQPEYVNPDPVRPP-----PSPRE-----GULP-AARPAQTLEPAKTL 1174
Db 1057 -----EYVNPQGETSRSLSDIYNPNVEDLTDGMPVSLSSQEAETNFSRPELT 1136
Qy 1175 FPKGVKVDVFAFGAVENPEYLTPOGAAPQHPHPPAFPMVLYWDDCPBERGAP 1234
Db 1107 TNQNSL---PLVSSGSMDDPY---QAG-----YQAAF-----LPOTGALT 1141
Qy 1235 STFKGTPTAENPEYLG 1250
Db 1142 GNGMFLPAENLEYLIG 1157

RESULT 9
ERBB3_HUMAN STANDARD; PRT; 1342 AA.
ID ERBB3_HUMAN
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Pinescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."

```

```

FT BINING 742 742 ATP (BY SIMILARITY)
FT ACT_SITE 834 834 BY SIMILARITY
FT DISULFID 186 194 BY SIMILARITY
FT DISULFID 190 202 BY SIMILARITY
FT DISULFID 210 218 BY SIMILARITY
FT DISULFID 224 226 BY SIMILARITY
FT DISULFID 227 235 BY SIMILARITY
FT DISULFID 231 243 BY SIMILARITY
FT DISULFID 246 255 BY SIMILARITY
FT DISULFID 259 286 BY SIMILARITY
FT DISULFID 290 301 BY SIMILARITY
FT DISULFID 305 320 BY SIMILARITY
FT DISULFID 323 327 BY SIMILARITY
FT DISULFID 500 509 BY SIMILARITY
FT DISULFID 504 517 BY SIMILARITY
FT DISULFID 520 529 BY SIMILARITY
FT DISULFID 533 549 BY SIMILARITY
FT DISULFID 552 565 BY SIMILARITY
FT DISULFID 556 573 BY SIMILARITY
FT DISULFID 576 585 BY SIMILARITY
FT DISULFID 589 610 BY SIMILARITY
FT DISULFID 613 621 BY SIMILARITY
FT DISULFID 617 629 BY SIMILARITY
FT CARBOHYD 126 126 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 250 250 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 353 353 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 408 408 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 414 414 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 437 437 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 469 469 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 522 522 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 566 566 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 616 616 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT VARSPLIC 141 183 ELSGGVTEKNDKCHMDTDMDIYADRABEIVKDNCR
SC -> GQFPMVPSGLTPQPMDDWTLDDDEPLTLISASSK
VPTLAIV (IN SHORT ISOFORM)
FT VARSPLIC 184 1342 MISSING (IN REF. 2)
FT CONFLICT 560 560 E -> G (IN REF. 2)
FT CONFLICT 1064 1064 E -> G (IN REF. 2)
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66C4374BD CRC64;

```

Query Match 34.0%; Score 2316.5; DB 1; Length 1342;
 Best Local Similarity 39.5%; Pred. No. 9.5e-117;
 Matches 520; Conservative 196; Mismatches 459; Indels 143; Gaps 36;

```

QY 10 GLLTALLPAGAA--STQVCTGTDMKRLPASPETHLDMLRLHYGCGVVOGNLELTLYPT 67
DB 11 GLPFLARGSEVNGSQAIVCPGTGLNGSTGDAENQYOTLYKLYICEVVMGNLELTLYTGH 70
QY 68 NASISFLQDIOEVQGYVLIANQVQVFLQRLIRVGTQLFEDNYVALAVLDNGDPLNNTT 127
DB 71 NADISFLQWIREVTVLVANNEFSTLPLMLRVYRGVYDGKFAIFVM-----LAVNT 125
QY 128 PVTGSFGLRELQRLSTELLKGVLIQRPOLCYOTILMKDIFHKGNOLATLTDN 187
DB 126 ----NSSHALQRLTQLTEILSGVYTEKNDKLCMDTIDMDIIVDRD---AEIYVKD 178
QY 188 RSRACHPCSPMCKSGRCWGESSEDDQSLRTIVCAGG-C-ARCKGFLPTDCHEOGCAAGCTG 246
DB 179 NGRSCPCHEVCKG-RCMGPGSEDDQTLTITICAPQCNCHGFCGNPNPOCCHDEACGCGSG 237
QY 247 PKHSDCLACLHPNHSIGTELHCPALVYNTDTFESMPNBEGRYFVGASCTTACPNYVLTST 306
DB 238 PQDDTDFACRHFHNSGACVPRCPQLVYVNTKLFQLEPFPHTKYGVGVGCVASCPHNFV-V 296
QY 307 DVSGCTVACPLHNOEVTAEQTOGCEKSKPCAVCGVGLGMQYIKANSKF--IGITILE- 363
DB 297 DQTCVTRACFPDKNHEVD-KNGIKNCEPFGGICCPKACBETG-----SGSRFTVDSNIDG 350
QY 364 FAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLOVEFTEITGTYVLIYISAMPDLPDLS 423
DB 351 FVNGTKILGNDFLITGLNGDPMHKIPALDPEKLVNFPRTVAIEITGTYVLIYISAMPDLPDLS 410

```

```

QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLISMLGRLSRELGSLALIHNTLHLCFVHTVPM 482
DB 411 VFSULTLTIIGRSYLNRRFSLIMKNNLVNTSLGFRSLKEISAGRIYISANQLCYHHSLSNW 470
QY 483 DQLFRRNHQALLHTA--NRPEDECVGEGLACHQICAGHCHMGPGPTQCVNCSQVLRQGEV 541
DB 471 TKVLRGFTERRLDIKHRRPRDCVABSKVCDPLCSSGCGPGCGQLSCRNSTRGVCV 530
QY 542 EECRVLOGLPREYVNAHCLPCHPECOPQNGSYTCFPEADQCVACAHYNDPFCVACRP 601
DB 531 THCNFLNGEPRFAHAEBCSCHPECQPMGTATCNCGSGSDTCAQCAHFFDGRHVCSSCP 590
QY 602 SGVPRDLSYPIPKFPEBAGCQPCPINCNSCVDDLDDCKCPAEQRA---SPLTSVSA 657
DB 591 HGVLG--AKSPIKYKPPVQNECRPCHENCTQGGCKGEBLQDCLQTLVLJCKTHLTMAVLT 648
QY 658 VGLGLLVVVLGVVFGILIKRROKIR-KYTRRLLOTEBLVEPLTSGAMPNOQMILK 716
DB 649 IAG--LVITFMGLGTLTYRGRRIQKRAMRYLERGESIEFLDPS-EKANKYLARIKF 705
QY 717 ETELRKXVVLGSAFGFNNFTVSFWLRVPKVSASH---LEVLRENTSPRANKELLIDEAY 772
DB 706 ETELRKXVVLGSAFGF---TVHKGVIPEGESIKIPVCIKVIEDKSGRQSPQAVTDHML 761
QY 773 VMAGVSPYVSRLIGTCLSTVQVLTQVLMFYGCLLVHRENRGRISQDILLNMCQATAG 832
DB 762 AIGSLDHAHIVRLIGLCPGSSLDLVQYPLGSLLDHVQRQALGPQLLLNNGVOIAGK 821
QY 833 MSYLEDVRLVHRDLAARVNVKSPNHKIDFGLABLDIDEVEYHADGKVPIKMALE 892
DB 822 MYLLEBHGWHRLAARVNLKSPQOVQVADPFGVADLLPPDDQQLYSEAKTIKMALE 881
QY 893 SIIRRRFTHOSDVSYSCTVWELMTFGAKPYDGI PARBEIPDLLEKGERLPQPICTIDVY 952
DB 882 SIHFQKXTHOSDVSYSCTVWELMTFGAEYAGRLAEVFDLLEKGERLQPOICTIDVY 941
QY 953 MIWVKCMIDSECRPPEELVSEFSMARPPQCFVYIYONEDLQRA---SPLDSTFRSLT 1009
DB 942 MWVVKCMIDENIRPPTFKELANFTTMARDPRLVLYIKES-GRGLAPGEPHGLTNKKL 1000
QY 1010 EDDMDGLVDABEYLVPOQGFPCPDPAFGAGVWHRHRSSTRSGGDLTGLERP-SEE 1068
DB 1001 EEVELBEPLDLDLDEAED-----NLATTTGSLSLVGLTLNRP 1041
QY 1069 EAPRSLPABEGAGSDVFDGDLGMAKGLQSLPTPD-ESPLORYSEDPVPLP----- 1121
DB 1042 RGSQSLSPSSGY-MPMNQNLGEGSCQESAVSSSERCPVSLH-----PMPROGLAS 1094
QY 1122 SETDGYVA-----PLTCSPOPE-----YVNOQPVPRQPPSPREGP----- 1156
DB 1095 ESSEGHVYTGSEALQEKVSMCRSRSSRSPRPGDSAYHSQRRSLTLPVTPLSPGLBEE 1154
QY 1157 -----LPAARPAATLERAKTLP-GKNGV-----KDVFAFGAVENPEYLTPOGGA 1203
DB 1155 DVNGVYVPTDHLKGTSSREGTSLVGLSLVGLTEDEED-----EEYEMNRRRH 1206
QY 1204 APOPHRPPAFSPAFDNLVYWD-----QDPPEKGAPESTFKGPTIANPEYL 1249
DB 1207 SP-PHPPRESSLELGEIYMDVGSGLSASLSTGSCPLHPVIMPTAGTTPDEVDYEM 1263

```

RESULT 10
 ERB3 RAT
 ID ERB3 RAT STANDARD; PRT; 1339 AA.
 AC 062759; 062955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (c-erbB3).
 GN ERB3
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA "Heller N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 cDNA and characterization of the
 RL recombinant protein.";
 RN Gene 165:279-284(1995).
 RP REVISIONS TO 85, 513 AND 565.
 RA Heller N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9010624;
 RA Carroll S.L., Miller M.L., Fromert P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroligin and their putative receptors, ErbB2 and
 RL ErbB3, is induced during Wallerian degeneration.";
 J. Neurosci. 17:1642-1659(1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTRK
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U29339; AAC28498.2; -
 DR EMBL: U52530; AAC53050.1; -
 DR HSPB; P11362; IFCX.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom: PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00219; TYKIC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP_FALSE_NEG.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRASNEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT BY SIMILARITY.
 FT BY SIMILARITY.

FT DISULFID 190 202 BY SIMILARITY.
 FT DISULFID 210 218 BY SIMILARITY.
 FT DISULFID 214 226 BY SIMILARITY.
 FT DISULFID 227 235 BY SIMILARITY.
 FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 629 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 FT SEQUENCE 1339 AA; 147545 MW; 0AASF2402BBDFDIE CRC64;
 SQ
 Query Match 32.9%; Score 2244.5; DB 1; Length 1339;
 Best Local Similarity 39.7%; Pred. No. 6.8e-113;
 Matches 512; Conservative 173; Mismatches 437; Indels 167; Gaps 37;
 3 LALCRGGLLALPPGAA---STQVCTGDMKRLPASPEHIDMKRLHYQCGQVQSN 59
 7 LQVLC---FLSLARGSEMSQAVCPGLTNGSLVTDADNOYQTLKYLKCEVYMGN 62
 60 LELVLPNLSLSDIOIGVGVVIAHNOVQVPLQRIYRGQLFEDNVALAVLDN 119
 63 LEIVLTGHNDLSFLQIRVYTVLVANNEFSVLPNLKRVRGTVQYQKRAIFVM-- 120
 120 GDPPLNTTPYTGASPGGLRELOLSLTELKGVLTORNPOLCYODITLKKDIFHNK 179
 121 ---LNTNT---NSSHLRQLKFTQLTEILSGVYIEKNDKLCIMDTIDRDIYRV-- 170
 180 ALTLIDNRSRACHPCGPMCKSGRCWGESEDDQSLTRYCAGGC-ARCKGRLPTCC 238
 171 GAETVVRNKGANCPGCHVECKG-RCWPGPDDQILTKTICAPQCNGRCGPPNQCCH 229
 239 QCAAGCTGPRGSDCLAHFNHSGICELHCPALVTYNTDPEESMPNPEGRYTGASCV 298
 230 ECAGGCGSGPDTCFACRRRNDGACVCPREPLVYNKLTFLQLEPNNHTYQGGCVAS 289
 299 CPNNYSLTDVGSCTLVCPPLNNGVTLAEDGTORCEKSKPCARVYGL--GMQYIKANSKF 356
 290 CPNHFV-VDOTFCVACPPDKMEVD-KHGILKMEPCGGGLCPKACEGSGSSRYQTVDS 347
 357 IGTLELFACCKTIFGSLAFSPGDDPSNAPLQPELOVFELELTGYLYISAWP 416
 348 ID----GFVVCITLGNLDLITGLVNDPNNKPLPALDEKLVNFRVREITGLNIO 403
 417 DSLPDLVSFQNLGVIRGRIILHNGAYS-LTLOGISMLGLSLRELSGLALIHNTHL 475
 404 PHMNFVSFNLTLTIGRSLYNGFSLLIKHNLNLTSLGPRSLKEISAGGVYISANQ 463
 476 FVHTVPPDDQLFRPHQALHTA-NREDEVEVGSLACHQLCARGHGWGSPQCVAC 534
 464 YHSLWMTKRLRPSSEERLDIKYDRPLGEGLAGKVCPLCGSGGCGWGPQGLSCR 523
 535 LRQGEVCEBRCVQLGQLPREYVNAHCLPGHPECOPNGSVTCGPGPADCVACAHYK 594

Db 524 SEGVGVCTHCHNCFQEPREFVHBAQCFSCHEPCELTMEGTSTCGSGSDACACAFRQGP 583
 QY 595 FCVARGSGVXKPDLSYMPIMKPPDEGACOPEDINTHSC--VDLDDKCKPEAGASPLT 652
 Db 584 HCNNSCPHGLG--AKGPIYKIPDAONECRPHENCTGCGNPELDDCLGQGLVMSKPH 641
 QY 653 STVSAYVGLLVVLGVFGLIKRQOKIR-KYTRRLLOETELVEPLTPSGAMPNQAQ 711
 Db 642 LVIAVTVG--LAVILMLGSGFLYWGRIQNRARRYLERGESIEPLDPS-ERANKVL 698
 QY 712 MLKETELRKXVVGSGAFGPNFVSWLVPKYSASH-----LEVRKNSPANKKEI 767
 Db 699 ARFTKETELRKLVLSGVFG---TVHKGIWPEGESIKIPVCIKVLEDKSGRSPQAV 754
 QY 768 LDEAYVAGVSGPYVRLGLICTSTVQLVTOLMPYGLCLDHVRENGRLSGQDLNNMC 827
 Db 755 TDMMLAVGLDHAHIVRLGLCGSSIQVLTQYLPGLSLDHVKQRETLGQDLNNMCV 814
 QY 828 QIAKMSYLEDVRLVHEDLAANNVLYKSPNNHYKITDGLARLLDDETEYHADGKVPK 887
 Db 815 QIAKMSYLEDVRLVHEDLAANNVLYKSPNNHYKITDGLARLLDDETEYHADGKVPK 874
 QY 888 WMALESILRRFTHSGDWSYGYTWELMTFGAKPYDGAPAREIPDLLEKGERLPQPPIC 947
 Db 875 WMALESILRRFTHSGDWSYGYTWELMTFGAKPYDGAPAREIPDLLEKGERLPQPPIC 934
 QY 948 TTDVYIMWYKMWIDSECPREFELVSEFSRMAPDQRFVVIQNEIDLGPASPLDSTFYRS 1007
 Db 935 TTDVYIMWYKMWIDSECPREFELVSEFSRMAPDQRFVVIQNEIDLGPASPLDSTFYRS 991
 QY 1008 LLEDMDGLVDAEELVYVQGFCDPAPAGAGMWHHRSSSTRSGGDLITLGEPE 1067
 Db 992 VLTTEL-----QEALEPEL-----DLDLLEAE 1017
 QY 1068 E-----EAPRSLAPSEG-----AGSVDFDCLMGAKIQ 1099
 Db 1018 EGLATSLGSLALPTLTLPRLGSGSLSSGYPMMNQSLSGEACLDVAVSGEQRFR 1077
 QY 1100 SLEPHDPSPLQRYSEDFVPLPSETDGYV---APL-----TC-----SQPE---Y 1139
 Db 1078 PLSLH-PIPRGR-----PASESSSEGHVTSGEAELOEKVYSYCRSRSRSPRPGDSAY 1129
 QY 1140 VNQPDVPPQPSPPREGE-----LPAPRAGATLERAKTISP-GKNQV----- 1181
 Db 1130 HQGRSLTLPTVPLSPGLEBEDNGYVMPDTHLKGASSREGTSSVGLSSVLTGERD 1189
 QY 1182 KDVFARGAIVENPEYLTPOGGAAPQPHPP 1210
 Db 1190 ED-----EEYEWNRKRQSP-PRPP 1209
 RESULT 11
 ID EGFR_DROME STANDARD; PRT; 1426 AA.
 AC P04412; O61601; Q9W2G0; P81868;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (EGFR)
 DE (Gutten receptor) (Torpedo protein) (Drosophila relative of ERBB)
 GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Preyoptera; Neoptera; Endopterygota; Diptera; Braachyera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7277;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
 RA MEDLINE=94350209; PubMed=8070664;
 RA Clifford R., Schubach T.;
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
 that several genetically defined classes of alleles cluster in

RT subdomains of the receptor protein.";
 RL Genetics 137:531-550(1994).
 RN [2]
 RP REVISIONS.
 RL Clifford R., Schubach T.;
 RN Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RA MEDLINE=85124611; PubMed=2982499;
 RA Lythen E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
 RT "The Drosophila EGF receptor gene homolog: conservation of both
 RT hormone binding and kinase domains.";
 RL Cell 40:599-607(1985).
 RN [4]
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
 RC STRAIN=Oregon-R; TISSUE=Embryo;
 RX MEDLINE=87002474; PubMed=3093080;
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
 RT "Alternative 5' exons and tissue-specific expression of the
 RT Drosophila EGF receptor homolog transcripts.";
 RL Cell 46:1091-1101(1986).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
 RP ANALYSIS.
 RA MEDLINE=99102120; PubMed=9882502;
 RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
 RT "Several levels of EGF receptor signaling during photoreceptor
 RT specification in wild-type, Ellipse, and null mutant Drosophila.";
 RL Dev. Biol. 205:129-144(1999).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hinkins R.A., Gale R.P.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,
 RA Abell R.F., Agbayani A., An H.-J., Andrews-Fiankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berns P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Butts K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson X., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moppherson D.,
 RA Melnikov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Schejter F., Shen H.,
 RA Shue B.C., Sidon-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Svrtek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [7]
 RP SEQUENCE OF 959-1078 FROM N.A.

CC STRAIN-Daekwanryeong;
 CC MEDLINE=85137938; PubMed=2963222;
 CC Wadsworth S.C., Vincent W.S. III, Bildeau-Wentworth D.;
 CC "A Drosophila genomic sequence with homology to human epidermal
 CC growth factor receptor.";
 CC Nature 314:178-180(1985).
 CC [8]
 CC SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
 CC ANALYSIS.
 CC MEDLINE=92038942; PubMed=1936959;
 CC Raz E., Schejter E.D., Shilo B.Z.;
 CC "interallelic complementation among DER/fib alleles: implications for
 CC the mechanism of signal transduction by receptor-tyrosine kinases";
 CC Genetics 129:191-201(1991).
 CC [9]
 CC REVIEW.
 CC MEDLINE=97248481; PubMed=9094709;
 CC Perimon N., Perkins L.A.;
 CC "There must be 50 ways to rule the signal: the case of the Drosophila
 CC EGF receptor.";
 CC Cell 89:13-16(1997).
 CC - FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-
 CC MAPK PATHWAY. INVOLVED IN A WIDED OF DEVELOPMENTAL DECISIONS.
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
 CC POLARITIES OF THE OCYCTE. IN THE EMBRYO, PLAYS A ROLE IN THE
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMIOSEOSA
 CC AND VENTRAL NEURODERMAL CELLS. GERM BAND RETRACTION, CELL FATE
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
 CC CUTICLE.
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC - SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
 CC PROTEIN.
 CC - ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: UNICITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLGEN OF
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
 CC AND THOACIC AND ABDOMINAL GANGLIA.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC - This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF052754; AAC08536.1; -;
 CC EMBL: AF052753; AAC08536.1; JOINED.
 CC EMBL: AF052754; AAC08535.1; -;
 CC EMBL: AF052752; AAC08535.1; JOINED.
 CC EMBL: K03054; AAAS1462.1; -;
 CC EMBL: K03417; AAAS1460.1; -;
 CC EMBL: K03416; AAAS0965.1; -;
 CC EMBL: K03418; AAAS1461.1; -;
 CC EMBL: AF109077; AAD26134.1; -;
 CC EMBL: AF109078; AAD26132.1; -;
 CC EMBL: AF109082; AAD26132.1; JOINED.
 CC EMBL: AF109078; AAD26133.1; -;
 CC EMBL: AF109084; AAD26133.1; JOINED.
 CC EMBL: AF109079; AAD26130.1; -;
 CC EMBL: AF109081; AAD26130.1; JOINED.
 CC EMBL: AF109079; AAD26131.1; -;
 CC EMBL: AF109083; AAD26131.1; JOINED.
 CC EMBL: AF109080; AAD26135.1; -;

DR EMBL: AB003454; AAF46732.1; -;
 DR EMBL: X02293; CAA26157.1; -;
 DR EMBL: X78820; CAA55523.1; -;
 DR EMBL: X78918; CAA55521.1; -;
 DR EMBL: X78919; CAA55522.1; -;
 DR PIR: A00640; GQFE.
 DR HSSP: P11362; IFGK.
 DR FlyBase: FB00003731; Egfr.
 DR InterPro: IPR000494; EGF_L_domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR Pfam: PF00069; Pkinase; 1.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR ProDom: PD000001; Euk_Pkinase; 1.
 DR SMART: SM00261; FU; 7.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE: PS00111; PROTEIN KINASE DOM; 1.
 KM Transmembrane, glycoprotein, Receptor, Phosphorylation; Transferase;
 KM Tyrosine-protein kinase: ATP-binding; Signal; Alternative splicing.
 KM Developmental protein.
 FT SIGNAL 1 30
 FT CHAIN 1 1426
 FT DOMAIN 31 868
 FT TRANSMEM 869 889
 FT DOMAIN 890 1426
 FT DOMAIN 938 1198
 FT NP_BIND 944 952
 FT BINDING 971 971
 FT ACT_SITE 1063 1063
 FT MOD_RES 902 902
 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
 Query Match 26.9%; Score 1836; DB 1; Length 1426;
 Best Local Similarity 31.8%; Pred. No. 5.5e-91;
 Matches 456; Conservative 183; Mismatches 436; Indels 360; Gaps 46;
 QY 24 QVCTGTDPKRLPASPETHLMDLRLHYOGGVVGNLELYLP-NASLSPLODIOEVOG 82
 DB 100 KICIGTSRLSPVSNKHHYANLDRNTCTCYVDGNLKTMLPHENDLSTLDNIREVTG 159
 QY 83 YVLIANQVRQVPLQRLRIYRQLF-----EDNYALAVLDNGDPLNNTPTVTGASPGGL 137
 DB 160 YVLIASHVDVKKVPPKQLIIRGRTLFSVSEEXYALFV-----TYSKM 203
 QY 138 RELQRLSTELTXGVLQIRNPOLCYODTILMKDIFKNNQALTLIDTNRBRACHPESP 197
 DB 204 YVLEIPDLRDVNLQVGFHNHYNLCHMRKTIQWSEIVSNGDAYNYDFTAPRECPKCHE 263
 QY 198 MCKSRGCVGSESSDCQSLTRTVGAGGA--RCKGPLTDCCHQCAAGCTGPKXSDCLAC 255
 DB 264 SETHG-CWGEGRPKCQKFKSKLTGSPQACAGRCYCPKRECHLFCAGGCTPTQKDIAC 322
 QY 256 LHPNHSGLCELHCAALVTYNTDTEFSNPBGRKTFGASCTATCPYNYLSTDVSCITIVC 315
 DB 323 KNFFDEAVSKKECPKRYNPTVLTNPBGKAYGATCKECP-CHLLADNGACVASC 381
 QY 316 PLHNQEVATBEDGTORCEKSKPCARVCYGLQMYIKANSKFIQITEL-----FFACCKK 369
 DB 382 PQDKMDKGE-----CVPNGFCPKTC-----PGTVLHAGNIDFRNCTV 422
 QY 370 IFGLSLFLPESFDG--DPASNTA-----PLQEPQLOVFTLEETITGLYISAMPDPLD 421
 DB 423 IDGINRIIDCTFFSFOVYANYTGPRIPLDPERRRRVFSVKEITGLNIEGTHPOFRN 482
 QY 422 LSVFQNLQVIRGRLHNGAY-SLTLOGIGSWLGLRLSRLREGSLALIHNNTHLCPHTV 480
 DB 483 LSYFRNLFTIGRQLMESFAALIVASSLSYSLMKRYLKOISSASVVIQHRDLCTYSNI 542
 QY 481 FMDQLFRNPHQALLHTANRPEDECVGGLACHQLCARGHGCPGPTQCVNCSQFLRQEC 540

```

Db      543 RPAIQLKEEKKYMWVENLRADLCENKNGTICSDQCNEDGCMWAGAGTQDCLTCNENFNNGTC 602
Qy      541 VEECVNLQGLPREYVNAHCLPCHPECCQNGNSVTCFGEADQCAVAHYXDPFECVARC 600
Db      603 IADCGYISNAK--FDNRICKICHPECR-----TCNGAGDHCOECVHRDGGCHVSEC 654
Qy      601 P-----SGVK-----PDL 608
Db      655 PKKRYNDRGVRECHATCDGCTGPKDTIGACTTCNLAINNDAIVKRLCKDKDCPD- 713
Qy      609 SYMPIWK--PDEEGACQ-----CPI-----NCTH----- 632
Db      714 GY--FWYVHPQGGSLKPLAGRAVCRKCHPLCELCCTVNGHYGYSCKTHYKREQCET 771
Qy      633 -----SC-----VDLDKG----- 641
Db      772 ECPADHYTDEBQRECFQRHPECGCTGPGADCKCRNPKLPDANETGTVNSTMNCTG 831
Qy      642 -CPAECR-----ASPLTS-----IVSAVGLLVVLGVVPGI 673
Db      832 KCPLEMRHVNVOYTAIGPYCAASPPRSSKITANLDVNMIFITGAVLVFTICILCV--T 889
Qy      674 LKBPQCKIRKXT--MRRLQETELVEPTPSGAMPNQAKMLKETEELRKXVLSGAF 731
Db      890 YICRQKQKAKKETVKTMTALSGCEDSEPLRPSNIGANLCKLIVDAELRKGCVLGMGAF 949
Qy      732 GNNFTVSPFMLRVP-----KVASHLVLENTSPKANKELIDEAYVAVGSPYVSRL 786
Db      950 G--RVYKGWV--VPEGENVKIPVAIKELK--STGASSESEF.REAYIVASEHVNLLKLL 1004
Qy      787 GICLTSTVQVLTQMPYGLLDHVENRRLSGSODLIMQMLAGMSTLEVLRLVHRDL 846
Db      1005 AVCMSSQMLITQLMPLGCLLDYVRNRRKIGSKALNNSTIAGMSTLEKRLVHADL 1064
Qy      847 AARNVLKSPNHVKTITDFGLARLLDIDETEHADGKVPYKMMALSLIRRFTHQSDVM 906
Db      1065 AARNVLVOTRPLVKITDFGLAKLSDSNHYKAGKMPKIMLAECIRNRFTSKSDVM 1124
Qy      907 SGTGVVWELMTFGAKPYDGIIPAREIPDLLEKEERLPPEPITTDVYIMVVKCMIDSECR 966
Db      1125 AGGVITWELTFGCPHENIPAKDIPDLLEVGLKEQPEICGLDIYCTLSGMDHAMR 1184
Qy      967 PPRFELVSEFMRARDPQRFVYIQNEDLG--PASPLDSFFVYSLLED--DMGVLVAE 1021
Db      1185 PTFKQLTTFAPAPDPGRKYLIPGDKFTRLA-----YTSQEKDLIRKLAFTTGS 1237
Qy      1022 EYLVPOQGFCDPAPAGAGMYHRRSSSTRSGGDLTLGLPSEEEAP-----RSPL 1075
Db      1238 EAIAPKPDYLOPKAAGPS-----HRTDCT-----DEMPKLNRYCKDPS 1276
Qy      1076 AASEGAGSVFPG--DLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLT 1132
Db      1277 NKNSSTGDDERSSAREVGVNLR-----LDLPVDDDDLMP--T 1314
Qy      1133 CSPOPEYVQDPVRPQPPRREGPLPAPAPAGATLERAKTISPKNKGVKQVAFAGAVE 1192
Db      1315 CGPGRNNNNNM-----NPNQNMMAAVGAVAGM-----DLIGPVAVD 1353
Qy      1193 NPEYL--TPQGAAPPH-----PPAFSP-AFDNLYYMD 1224
Db      1354 NPEYLLMAQTLGVGESPIPTQIGIPVWGPGTMEVYKVPMPGSEFTSSDHEYYND 1408

```

```

RESULT 12
ERBB_ALV
ID_ERBB_ALV STANDARD: PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
V-ERBB.

```

```

OS Avian leukosis virus.
CC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
CX NCBI_Taxid=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8528222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIA ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
CC EMBL, M10066; AAA48763.1; ALT_INIT.
CC PIR, A00643; TVCHLV.
CC PIR, B00643; TVFVLV.
CC HSSP, P11362; 1FGK.
CC InterPro: IPR000719; Euk PKinase.
CC InterPro: IPR001245; Tyr PKinase.
CC Pfam, PF00069; Kinase; 1.
CC PRINTS: PR00109; TYRKINASE.
CC PRODOM: PD000001; Euk PKinase; 1.
CC SMART, SM00219; Tyrc; 1.
CC PROSITE, PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE, PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE, PS50011; PROTEIN KINASE_DOM; 1.
CC Translatae; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399
CC NP BIND 138 146 ATP (BY SIMILARITY).
CC BINDING 165 165 ATP (BY SIMILARITY).
CC ACT SITE 257 257 BY SIMILARITY.
CC FT ACT SITE 257
CC SQ SEQUENCE 634 AA; 70891 MW; E705533A0BE01FCC CRC64;

```

```

Query Match 24.0%; Score 1638.5; DB 1; Length 634;
Best Local Similarity 50.2%; Pred. No. 8.36-81;
Matches 357; Conservative 80; Mismatches 143; Indels 131; Gaps 19;

```

```

Qy 587 CAHYKDDPPCYARPPSSVGRKDLSTMPYWKPPDEGAQCPINPINTHSCVLDLDDGCPAQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VKXADANAVQQLCHPNTCRCKGPGLEGCP-- 58
Qy 647 RASPLTISVAVV-GILLVVLGVVFGILLKRRQCKIRKTYMRLLQETELVEPLTPSGA 705
Db 59 NGSKTPISAGVVGGLCLVVGIGLILYRR-HYRKXTRRLLOERLEVEPLTPSGE 117
Qy 706 MPNOQMRILKETELRKVYKVLGSGAGFNNFTVSWLVRPKVASHLV--LRENTSP 761
Db 118 APNOAHRLIKETELFFKVKVYLGSGAFG---TYKKGIMDEGEKVKIPVAIKELRENTSP 173
Qy 762 KANKEIIDEAYVAVGSPYVSRLGLCLSTVQVLTQMLPYGCLLDHVENRRLSGSD 821
Db 174 KANKEIIDEAYVAVSUNPHVCRLLGLCLSTVQVLTQMLPYGCLLDHVENRRLSGSD 233
Qy 822 LLNMCQIAGMSTLEVDRLVHDLAARNVLYKSPNHVKTITDFGLARLLDIDETEHADG 881

```

```

DB 234 LNMVQYAKGMNVLLEERLVHRDLAARVILVKTQHVKITDFTGLAKLGNDEKRYHEG 293
QY 882 GYVPIKMALESILRRRFTHQSDWSYGVVWELMTFGAKKPDIGIPABEIPDLKEGERL 941
DB 294 GYVPIKMALESILRRRFTHQSDWSYGVVWELMTFGSKPYDGIPIASISSVLEKGERL 353
QY 942 POPPCTIDVYIMVYCMVISECRPFRELSERSRMARDQRFVVIQ-NEDLGPAAPL 1000
DB 354 POPPCTIDVYIMVYCMVISECRPFRELSERSRMARDQRFVVIQ-NEDLGPAAPL 413
QY 1001 DSFTFRSLLEDDMGDLVDAEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLT 1060
DB 414 DSFTFRSLLEDDMGDLVDAEYLVPOQGFPCPDPAAGAGVHHRRSSSTRSGGDLT 449
QY 1061 LGLEPSEERAPSPL-----APSEGSGVDFPDGDLGMGAKGLOSLPTDPSPLQRYSED 1115
DB 450 -----SKPLISLSATSNNSATNCID-----RNGQGHVREDSEFVQRYSSD 491
QY 1116 PTVPLPSEF--DGYVAPLTCSPQPEYVNPDPVPPSPBEGPLPAAPAGATLERATL 1173
DB 492 PTVPLPSEF--DGYVAPLTCSPQPEYVNPDPVPPSPBEGPLPAAPAGATLERATL 522
QY 1174 SPQKGVVAVDF-----AFGAVENPEVITTPQGAAPQHPAPSPAPD 1218
DB 523 AVVQVQIYNNISLTAISKLPMDSRVONSHSTAVNPEYL-----NTNQSPLAKTYFE 574
QY 1219 NUYVMDQ-----DPE-----RGAPSTFGKTPAENPEYGLDVP 1254
DB 575 SSFYWIQSGNHQINLNDPNPYQDFLENKRPNGLLKVPAAENPEYLRYAAP 625

RESULT 13
EGFR CHICK STANDARD; PRT; 703 AA.
AC P13387;
AD 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 1;
RX SEQUENCE FROM N.A. PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ulrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Biol. 8:1970-1978 (1988).
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB 8 EMBL; M20386; AAA48760.1; -
DB InterPro; IPR000494; EGFR_L_domain.
DB InterPro; IPR000719; EGF_pkinase.
DB InterPro; IPR002174; Furin-like.
DB InterPro; IPR001245; Tyr_pkinase.
DB Pfam; PF00757; Furin-like; 1.
DB Pfam; PF00757; Furin-like; 2.
DB Pfam; PF00757; Furin-like; 3.
DB Pfam; PF00757; Furin-like; 4.
DB PROSITE; PS00107; PROTEIN KINASE_TYR; PARTIAL.
DB PROSITE; PS00109; PROTEIN KINASE_TYR; PARTIAL.
DB PROSITE; PS00111; PROTEIN KINASE_DOM; PARTIAL.
DB Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; phosphorylation.
FT SIGNAL 1..30
FT CHAIN 31..703
FT DOMAIN 31..654
FT TRANSMEM 655..667
FT DOMAIN 668..703
FT DISULFID 197..206
FT DISULFID 201..214
FT DISULFID 222..230
FT DISULFID 226..238
FT DISULFID 239..247
FT DISULFID 243..255
FT DISULFID 258..267
FT DISULFID 271..298
FT DISULFID 302..314
FT DISULFID 318..333
FT DISULFID 336..340
FT DISULFID 340..344
FT DISULFID 513..522
FT DISULFID 517..530
FT DISULFID 533..542
FT DISULFID 546..562
FT DISULFID 565..581
FT DISULFID 569..589
FT DISULFID 592..601
FT DISULFID 605..627
FT DISULFID 630..638
FT DISULFID 634..646
FT CARBOHYD 134..134
FT CARBOHYD 190..190
FT CARBOHYD 200..200
FT CARBOHYD 359..359
FT CARBOHYD 368..368
FT CARBOHYD 420..420
FT CARBOHYD 573..573
FT CARBOHYD 578..578
FT CARBOHYD 613..613
FT CARBOHYD 633..633
FT CARBOHYD 648..648
FT NON_TER 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE1B735A690 CRC64;

Query Match 23.4%; Score 1595; DB 1; Length 703;
Best Local Similarity 44.8%; Pred. No. 2e-78;
Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;
-----
QY 8 RFWGLLALPPGAA-----STGYCTGDMKRLPASPETHLDMRLHYGCGVYQGNLE 61
DB 13 RGAVALVLLLVGLVAGLCAVEBKVCQGTNNKLLQLGHEVHFSLQRMVNNCEVLSNLE 72
QY 62 LTYPTASLSPFLDIOEVGYVLIANQVRPPLQRLRVRTQLPEDNYALAVLNGD 121
DB 73 ITVEHNRDLTFKTTQDEVAGVLIANMVDVPLELQIRGVLYVDSNFALAVSNVH 132
QY 122 PLNNTVTGASPCGLRELDLRSTELTKGVYIQRPOCYOPTIIMKPIFKNNQAL 181
DB 133 -MNTQ-----GLRELPMKLSLTLNGVATSNPKLCNMDVYAMNDIITSKR-PL 182
QY 182 TLID-TNRSPACHPCSMCKGSRGWSESSDCSLRTVYAGGCA-RCKGPLEPTDCHEQ 239

```

```

DB 183 TVLDFASNLSSCPKCHNCTEDHOCWAGBONCOTLTKVICAQCCSRGKVPSCCHNQ
QY 240 CAAGCTGPHSDCLACLFHNSGICELHPALVATYTDTPESMPNDEGGTYGASCVTAC
DB 243 CAGAGCTGPRSDCLACRKRFRDQATCQDTPPLVLYPTTYQMDVNPDEGRTSGATCVRBC
QY 300 PNYVLTSDVSGCTLVCPHLNQEVTAEQDGTGCEKSKPCARVCGYLGWGYIKANSKFTGI
DB 303 PHNYVATDHGSCVRSQNTDTYEV-EENGVKCKCKCGGLSKVNCNGIGELGILS-INA
QY 360 TELE-PAGCKKIFGSLAFPESEDPDPAQNTAPLOEQOVETLEIIGLYIYIAMPDS
DB 361 TNDSEKNCCTKINGDVSLPVALGAFKTLPLDPKLDVFRVYAEISGFLLIQAMPBN
QY 419 LPLDSYFQNLQVIRGRIHNGAVSLTLQGLISWLGSLRSRELSGSLAIHNHTLCEFH
DB 421 ATDLVAFENLEIRGTRKQHQGSLAVNLKIQSLGRLKEISDDIIMKNKULCVAD
QY 479 TVWMDOLFRRPHOALHTANRPDECEVGGGLACHQICAGHCKWGPPTQCVNCSOFLRQ
DB 481 TMMRSLFATQSGTKIKIONRKNDCDADRHYCDPLGSDVCGWGPFPFSCRFSSRQK
QY 539 ECVEECRVGLQGLPREYVNAHCLPCHPECOPQNG--SYTCFGEPAQCVCAHYKDPF
DB 541 ECVKQCNIIQGEPRERERDSKCLPCHSECLVQNSTAYNTGSGPGRDHCKMAHFDGH
QY 596 CVARCPGVKPDLSYPIWKFPEDEGACQPCPINTCHSCVDDDKCPAEGASPLTSIV
DB 601 CVACAPAGVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGFLGECF--NGSKTPSIA
QY 656 SAVV-GILIVVLGVVFGILIKRROOKIRKTYMRRLLOTELEVEPLTP
DB 657 AGVVGSLCLVVGGLGILYLRNR-HIVKRTILRLLORELEVEPLTP

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
DB EMBL, K02006; AAA42394.1; ALT INIT.
DB EMBL, K01216; AAA42400.1; -.
DB PIR, A00644; TYVTH.
DB HSSP, P11362; 1BEK.
DB InterPro, IPR000719; Euk_kinase.
DB InterPro, IPR001245; Tyr_kinase.
DB Pfam, PF00069; Pkinase.1.
DB ProDom, PD000001; Euk_Pkinase; 1.
DB SMART, SMC0219; TyrKc; 1.
DB PROSITE, PS00107; PROTEIN_KINASE_ATP; 1.
DB PROSITE, PS00109; PROTEIN_KINASE_TYR; 1.
DB PROSITE, PS00011; PROTEIN_KINASE_DOM; 1.
DB Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
DB Glycoprotein; Phosphorylation.
DB DOMAIN 132 399
DB NP_BIND 138 146 ATP (BY SIMILARITY).
DB BINDING 165 165 ATP (BY SIMILARITY).
DB ACT_SITE 257 257 BY SIMILARITY.
DB CONFLICT 29 29 R -> W (IN REF. 2).
DB CONFLICT 140 140 S -> F (IN REF. 2).
DB CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BCDD06745D609 CRC64;

Query Match 23.4%; Score 1592; DB 1; Length 604;
Best Local Similarity 50.0%; Pred. No. 2.4e-78;
Matches 347; Conservative 77; Mismatch 136; Indels 134; Gaps 18;

QY 587 CAHYKDPFCVACRCPGVKPDLSYPIWKFPEDEGACQPCPINTCHSCVDDDKCPAEG 646
DB 3 CAHFDGPHCVKACRAGVLGENDTL-VKRYADANAVCOLCHPNCTRGCKGFLGECF-- 58
QY 647 RASPLTSIVSAVV-GILIVVLGVVFGILIKRROOKIRKTYMRRLLOTELEVEPLTPSGA 705
DB 59 NGSKTPSIAAVVVGGLCLVVGGLGILYLRNR-HIVKRTILRLLORELEVEPLTPSGE 117
QY 706 MPNQAQMKLKETELRKVKVLGSGAFGNNFTVSFWLRYPVKASHLEV----IRENSP 761
DB 118 APNQHLKILKETEPKVKVLGSGAFG---TIKGLMTEGKVKVLAIKELREATSP 173
QY 762 KANKEIIDEAAYVAVGSPVYRLIGICTSTVOLVQLMPYGGCLLDHVRNRRGLSGOD 821
DB 174 KANKEIIDEAAYVAVSVDHPVCRLLGICLTSTVOLQLMPYGGCLLDYIREHKNIGSQY 233
QY 822 LLNMCQIAKGMSTLEDVRLVHRLAARVLYKSPNHYKTKDFFGLARLLDDETEYHAGD 881
DB 224 LLNMCVOIAKGMNLEERRLVHRLAARVLYKSPNHYKTKDFFGLARLLDDETEYHAGD 293
QY 882 GKVPDKMALESIRRRFTHSDVMSGVVWELMTGAKPYDGIAPAREIPDLLEKGRLL 941
DB 294 GKVPDKMALESIRHRYTHSDVMSGVVWELMTGAKPYDGIAPAREIPDLLEKGRLL 353
QY 942 POPPCTIDVYMWKCMIDSECRPRELIVSEFSRMAQDPQFVVIQ-NEDIGPASPL 1000
DB 354 POPPCTIDVYMWKCMIDADSRPRELIVSEFSRMAQDPQFVVIQ-OCGERNHLPSPT 413
QY 1001 DSTFYRLSDDDMDGLVDAEYVLPVQGFCCPRPARGAGAGMTHRRSSSTRGGGDLT 1060
DB 414 DSKFYRLSDDEDEDVDAEYVLPVQGF--NSPST----- 445
QY 1061 LGLPSEBEARSPF-----APSEGASVDFDGLGMAKGLSLPTPHSPLOQYSED 1115
DB 450 -----SRPLSLSLSTSNNSATNCID-----RNGGHVREDSEFYQRYSSD 491
QY 1116 PTVLPLPSET-DGVVAULTGSPQPEYVNPQVRQPSRREGPLPAARPAQATLERAKTL 1173

```

DB 492 PTGNFLEESIDDEFL-----PAPEVYNQ--LMPKXSTAM----- 524

QY 1174 SPKNGVYKVPFAF-----GGAIVENPEYTPQGAAPQPPHAPS 1214

DB 525 -----VQNOIYNFISLTAISKLPMDSRVYNSHSTAVDNPEYL-----NTNOSPLAK 570

QY 1215 PAEDNLYWDDPPERGAPSTFKGTPTAENPEY 1248

DB 571 TVESSPYWIOGNNHQ-----INLDNPDY 594

RESULT 15

ERBB_AVIEU STANDARD; PRT; 540 AA.

ID ERBB_AVIEU STANDARD; PRT; 540 AA.

AC P1273;

DT 01-JUL-1989 (Ref. 11, Created)

DT 01-JUL-1989 (Ref. 11, Last sequence update)

DT 15-JUN-2002 (Ref. 41, Last annotation update)

DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).

GN V-ERBB.

OS Avian erythroblastosis virus (strain ts167).

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.

OX NCBI_TaxID=103898;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=67064456; Pubmed=2878364;

RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;

RT "A single amino acid substitution in v-erbB confers a thermolabile phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells."

RT Mol. Cell. Biol. 6:1751-1759(1986).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: M13179; AAA42401.1; -

DR PIR: A25231; TVFVEB.

DR HSSP: P11362; 1FGK.

DR InterPro: IPR000719; Euk_pkinase.

DR InterPro: IPR001245; Tyr_pkinase.

DR Pfam: PF00069; pkinase; 1.

DR ProDom: PD000001; Euk_pkinase; 1.

DR SMART: SM00219; TyrKc; 1.

DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.

DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.

DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;

KW Glycoprotein; Phosphorylation.

FT DOMAIN 132 399

FT NP BIND 138 146

FT BINDING 165 165

FT ACT SITE 257 257

FT VARIANT 270 270

FT SEQUENCE 540 AA; 60412 MW; 5B53297A068B65D CRC64;

QY

Query Match 22.4%; Score 1524; DB 1; Length 540;

Best Local Similarity 52.8%; Pred. No. 9.3e-75;

Matches 329; Conservative 70; Mismatches 126; Indels 98; Gaps 16;

QY 587 CARYKDPFVACPSGVKDLVMPWKFPDEGACQPPINCHSVDLDDKGPAPQ 646

DB 3 CAHFIDGPHCVKACPGAVIGENDTL-VMKYADANAVCOJCHPNCRTGCKGPGLEGCP--- 58

QY 647 RASPLTISVAVV-GILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVPLTPSGA 705

DB 59 NSKTPSIAGVVGGLLCLVVVGLIGLYLRR-HIVRKRLRLLOERELVBLTPSGE 117

QY 706 MPNOQWMLKETELRKXKVLGSAFGNNFTVSFWLRVPKVSASHLEV-----LRENTSP 761

DB 118 APNOAHILKLETEKXKVLGSAFG---TVKGLMIPGEKVTTPVAIKELREATSP 173

QY 762 KANKEILDEAVYVAVGSPPVSRLLGLCLTSVQLVQLMFGCLLDHVRNRRGLSQD 821

DB 174 KANKEILDEAVYVAVSVDPNPHVCRLLGLCLTSVQLVQLMFGCLLDHVRNRRGLSQD 233

QY 822 LLNMCQIAKGNSTYEDVRLVHRDLAARNVLYKSPNRYKTDPEGLARLLDDELEYHAGS 881

DB 234 LLNMCQIAKGNSTYEDVRLVHRDLAARNVLYKSPNRYKTDPEGLARLLDDELEYHAGS 293

QY 882 GKVPKXWALSLRRRTTHQSDVWYGVTVWELMTFGAKPYDGIAPAREIDLEKGERL 941

DB 294 GKVPKXWALSLHRTYTHQSDVWYGVTVWELMTFGAKPYDGIAPAREIDLEKGERL 353

QY 942 POPPICTIDVYMIWYKCMWIDSECRPRERELVSFESRMAKDPQRFVYIQ-NEDUGPASP 1000

DB 354 POPPICTIDVYMIWYKCMWIDSECRPRERELVSFESRMAKDPQRFVYIQ-NEDUGPASP 413

QY 1001 DSTFYSRLLEDMDGDLDVDAEYLVPOGFCPPDPAQAGMWHRRSSSTRGGGLT 1060

DB 414 DSKFYRLMESEDMEDIVDADEYLVPOGFCPPDPAQAGMWHRRSSSTRGGGLT 449

QY 1061 LGLPSEEEAPRSPL-----APSEGAGSDVECDGLGMAKGLSLPTHPDPSPLQRYSED 1115

DB 450 -----SRTPLSLSLATSNNSATINCIDRNG-----H----- 476

QY 1116 PTVPLPSTDGVVAPLTSPPQPEYVNDQVAPQPPSPREGLPAARPAAGAT-LEBAKLS 1174

DB 477 -----PVREDGFL-----PAPEVYNQ--LMPKXSTAMVQNOIYNFISLTAISKLPIDS 523

QY 1175 PKNGVYKVPFAF-----GGAIVENPEYL 1197

DB 524 RYQN-----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:19:00

Job time : 21.642 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.8246 Seconds

(without alignments)
5522.503 Million cell updates/sec

Title: SEQ4-710-730-14

Perfect score: 6813
Sequence: 1 METALCRWGULLALLPPGA.....TFKGTPTANEPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6091	89.4	1259	6	018735
2	3033	44.5	1209	1	03QX70
3	3004	44.1	1210	11	03EP98
4	2617	38.4	1165	13	03YH40
5	2580.5	37.9	1137	13	03W666
6	2204	32.3	1328	13	P79754
7	1923.5	28.2	1433	5	09B1H9
8	1871	27.5	419	4	09UK79
9	1739	25.5	367	11	08R2X1
10	1697.5	24.9	412	15	08RYV0
11	1609	23.6	729	15	086712
12	1607	23.6	567	15	086714
13	1547.5	22.7	562	15	084895
14	1533	22.6	545	15	085468
15	1506.5	22.1	655	11	09WV55
16	1490.5	21.9	643	11	09ERV6

17	1201	17.6	1193	5	09Y1X8	09Y1X8 ephydactia f
18	1126	16.5	527	13	090836	090836 gallus gall
19	1117.5	16.4	1368	5	023821	023821 caenorhabdi
20	1114	16.4	1717	5	026566	026566 schistosoma
21	1001.5	14.7	478	11	09SE50	09SE50 rattus norv
22	942.5	13.8	599	13	09PSH2	09PSH2 gallus gall
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.0	176	11	0923V5	0923V5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	08S2M1	08S2M1 xiphophorus
27	754.5	11.1	311	13	099162	099162 xiphophorus
28	734	10.8	331	4	09BUD7	09BUD7 homo sapien
29	723	10.6	149	6	09BC66	09BC66 oryzocolagus
30	695.5	10.2	1362	13	09PVZ4	09PVZ4 xenopus lae
31	679	10.0	1671	5	09NVJ5	09NVJ5 biophthalari
32	655	9.6	1418	13	093457	093457 scophthalmu
33	646.5	9.5	1368	13	08UW85	08UW85 paralicthy
34	631.5	9.3	1369	13	08UW86	08UW86 paralicthy
35	630.5	9.3	1472	5	09U5A8	09U5A8 bombyx mori
36	621	9.1	1412	13	08UW84	08UW84 paralicthy
37	611	9.0	1358	13	073798	073798 xenopus lae
38	604.5	8.9	1418	13	08UW83	08UW83 paralicthy
39	592	8.7	1245	13	09YGH8	09YGH8 scophthalmu
40	583	8.6	1371	11	09QVW4	09QVW4 rattus sp.
41	580.5	8.5	2144	5	09VD94	09VD94 drosophila
42	578.5	8.5	1091	4	09UW04	09UW04 homo sapien
43	572	8.4	987	11	091YMO	091YMO mus musculu
44	569	8.4	935	4	096L35	096L35 homo sapien
45	569	8.4	987	11	099WR2	099WR2 mus musculu

ALIGNMENTS

RESULT 1

018735 PRELIMINARY; PRT; 1259 AA.

ID 018735

AC 018735

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ErpB-2.

OS Canis familiaris (Dog).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI_TaxId=9615;

RN 11

RP SEQUENCE FROM N.A.

RA Yokota H.;

RT "CDNA cloning of erpB-2 from canine mammary gland."

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB008451; BAA23127.1; ..

DR HSSP; P11362; ITGK.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR000719; Euk_Pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr_Pkinase.

DR InterPro; IPR004019; VLP_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; pkinase; 1.

DR Pfam; PF01030; Recep_L domain; 2.

DR Pfam; PF02757; VLP_2.

DR ProDom; PD000001; Euk_pkinase; 1.

DR SMART; SM00261; Fur_3.

DR SMART; SM00219; TyrcKc; 1.

DR PROSITE; PS00018; EF_HAND; UNKNOWN; 1.

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase.

SEQUENCE 1259 AA; 137969 MW; E37364D4C4CD46 CRC64;

Query Match 89.4%; Score 6091; DB 6; Length 1259;
 Best Local Similarity 89.2%; Pred. No. 0;
 Matches 1128; Conservative 49; Mismatches 72; Indels 16; Gaps 5;

QY 1 METALACRGLLALLPPGAASQTGTDMKLRLEPASETHLDMRLHYQCCQVYQGL 60
 DB 1 METALACRGLLALLPPGAASQTGTDMKLRLEPASETHLDMRLHYQCCQVYQGL 60
 QY 61 ELTYLPTNALSFLQDIQEVQVYLIAHQVQVPLQRLRYRGQLFEDNVALAVLNG 120
 DB 61 ELTYLPTNALSFLQDIQEVQVYLIAHQVQVPLQRLRYRGQLFEDNVALAVLNG 120
 QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRPDQCYDTIIMKDI FKHKNQLA 180
 DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRPDQCYDTIIMKDI FKHKNQLA 180
 QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRPDQCYDTIIMKDI FKHKNQLA 180
 DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLLIQRPDQCYDTIIMKDI FKHKNQLA 180
 QY 181 LTLIDTNRBACHPCSPMCKGSRGWSESESDQSLTRTCAGACGACRCKPPTDCCHEQC 240
 DB 181 LTLIDTNRBACHPCSPMCKGSRGWSESESDQSLTRTCAGACGACRCKPPTDCCHEQC 240
 QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDRESMPNDEGRYTFGASCVTACP 300
 DB 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDRESMPNDEGRYTFGASCVTACP 300
 QY 301 YNYLSTVSGCTLVGCLHNOETAEEDTQCEKSCRCARVCGLMQVYKANSKFIQT 360
 DB 301 YNYLSTVSGCTLVGCLHNOETAEEDTQCEKSCRCARVCGLMQVYKANSKFIQT 360
 QY 301 YNYLSTVSGCTLVGCLHNOETAEEDTQCEKSCRCARVCGLMQVYKANSKFIQT 360
 DB 301 YNYLSTVSGCTLVGCLHNOETAEEDTQCEKSCRCARVCGLMQVYKANSKFIQT 360
 QY 361 ELEFAGCKKIFGSLAFLPESFDGPASNTAPLOPEOLQVFEETLEITGLYISAMPDLP 420
 DB 361 ELEFAGCKKIFGSLAFLPESFDGPASNTAPLOPEOLQVFEETLEITGLYISAMPDLP 420
 QY 361 ELEFAGCKKIFGSLAFLPESFDGPASNTAPLOPEOLQVFEETLEITGLYISAMPDLP 420
 DB 361 ELEFAGCKKIFGSLAFLPESFDGPASNTAPLOPEOLQVFEETLEITGLYISAMPDLP 420
 QY 421 DLSVFONLOVIRGRIIANGAYSLTLQGLISWGLRSLRELCSGIALIHNTHLCFVHTV 480
 DB 421 DLSVFONLOVIRGRIIANGAYSLTLQGLISWGLRSLRELCSGIALIHNTHLCFVHTV 480
 QY 421 DLSVFONLOVIRGRIIANGAYSLTLQGLISWGLRSLRELCSGIALIHNTHLCFVHTV 480
 DB 421 DLSVFONLOVIRGRIIANGAYSLTLQGLISWGLRSLRELCSGIALIHNTHLCFVHTV 480
 QY 481 PMDOLFRNPHOALLHNRANPEDECEGELACHQLCARHGMCPGPTQCNCSQPLRGRC 540
 DB 481 PMDOLFRNPHOALLHNRANPEDECEGELACHQLCARHGMCPGPTQCNCSQPLRGRC 540
 QY 481 PMDOLFRNPHOALLHNRANPEDECEGELACHQLCARHGMCPGPTQCNCSQPLRGRC 540
 DB 481 PMDOLFRNPHOALLHNRANPEDECEGELACHQLCARHGMCPGPTQCNCSQPLRGRC 540
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQONGSVTCFGBEADQVACAAYKDPFCVARC 600
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECPQONGSVTCFGBEADQVACAAYKDPFCVARC 600
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQONGSVTCFGBEADQVACAAYKDPFCVARC 600
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECPQONGSVTCFGBEADQVACAAYKDPFCVARC 600
 QY 601 PSQVPRDLSYMPYKWRPEDEGACORPCINCTHSCVDLDKGGPACQASPVTSITAAVVG 659
 DB 601 PSQVPRDLSYMPYKWRPEDEGACORPCINCTHSCVDLDKGGPACQASPVTSITAAVVG 659
 QY 601 PSQVPRDLSYMPYKWRPEDEGACORPCINCTHSCVDLDKGGPACQASPVTSITAAVVG 659
 DB 601 PSQVPRDLSYMPYKWRPEDEGACORPCINCTHSCVDLDKGGPACQASPVTSITAAVVG 659
 QY 661 ILLVVLGVVFGILIKRQOKIRKYMRELLOETELVEPLTPSGAMPNQAQWRIKETEL 720
 DB 661 ILLVVLGVVFGILIKRQOKIRKYMRELLOETELVEPLTPSGAMPNQAQWRIKETEL 720
 QY 661 ILLVVLGVVFGILIKRQOKIRKYMRELLOETELVEPLTPSGAMPNQAQWRIKETEL 720
 DB 661 ILLVVLGVVFGILIKRQOKIRKYMRELLOETELVEPLTPSGAMPNQAQWRIKETEL 720
 QY 721 RKVKULGSGAFGNNFTVFWLRP----KVSASHLEVLRENTSPKANKELIDRAYYMA 775
 DB 721 RKVKULGSGAFGNNFTVFWLRP----KVSASHLEVLRENTSPKANKELIDRAYYMA 775
 QY 721 RKVKULGSGAFGNNFTVFWLRP----KVSASHLEVLRENTSPKANKELIDRAYYMA 775
 DB 721 RKVKULGSGAFGNNFTVFWLRP----KVSASHLEVLRENTSPKANKELIDRAYYMA 775
 QY 776 GVGSPVSRLLGLCTSTVOLTOIMPYGLLDHYRENGRSGSODLNMCMQIAKMSY 835
 DB 776 GVGSPVSRLLGLCTSTVOLTOIMPYGLLDHYRENGRSGSODLNMCMQIAKMSY 835
 QY 776 GVGSPVSRLLGLCTSTVOLTOIMPYGLLDHYRENGRSGSODLNMCMQIAKMSY 835
 DB 776 GVGSPVSRLLGLCTSTVOLTOIMPYGLLDHYRENGRSGSODLNMCMQIAKMSY 835
 QY 835 LEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPYKMALESIL 895
 DB 835 LEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPYKMALESIL 895
 QY 835 LEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPYKMALESIL 895
 DB 835 LEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPYKMALESIL 895
 QY 896 RRRFTHQSDVWSYGVTVVWELMTFGAKPYGIPARELPDLLEGEELPOPPCTIDVYIM 955
 DB 896 RRRFTHQSDVWSYGVTVVWELMTFGAKPYGIPARELPDLLEGEELPOPPCTIDVYIM 955
 QY 896 RRRFTHQSDVWSYGVTVVWELMTFGAKPYGIPARELPDLLEGEELPOPPCTIDVYIM 955
 DB 896 RRRFTHQSDVWSYGVTVVWELMTFGAKPYGIPARELPDLLEGEELPOPPCTIDVYIM 955
 QY 955 VECWMDISECRPRFRELVAEFSRMAADPQRFVYVIONEDLSPASPLDSTFYRSLLEDDMG 1015
 DB 955 VECWMDISECRPRFRELVAEFSRMAADPQRFVYVIONEDLSPASPLDSTFYRSLLEDDMG 1015
 QY 955 VECWMDISECRPRFRELVAEFSRMAADPQRFVYVIONEDLSPASPLDSTFYRSLLEDDMG 1015
 DB 955 VECWMDISECRPRFRELVAEFSRMAADPQRFVYVIONEDLSPASPLDSTFYRSLLEDDMG 1015
 QY 1016 DLVDAEVLVPOGFFCPDPAAGAGMTHHRSSSTRSGGDLTLGLEPSEEEAPRSPL 1075
 DB 1016 DLVDAEVLVPOGFFCPDPAAGAGMTHHRSSSTRSGGDLTLGLEPSEEEAPRSPL 1075

DB 1015 DLVDAEVLVPOGFFCPDPAAGAGMTHHRSSSTRSGGDLTLGLEPSEEEAPRSPL 1074
 QY 1076 APSEGASDVFDGGLGMAAKGLOSLTHHPSPLOKRSSEPTVPLSEETGCVAPLTCSP 1135
 DB 1075 APSEGASDVFDGGLGMAAKGLOSLTHHPSPLOKRSSEPTVPLSEETGCVAPLTCSP 1134
 QY 1136 QPEYVNPQDVARPPSPRESEPLPAARPAATLER-----AKTSPGKNGVYKDYFAFGA 1190
 DB 1135 QPEYVNPQDVARPPSPRESEPLPAARPAATLER-----AKTSPGKNGVYKDYFAFGA 1194
 QY 1191 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPERGAPSTKGTPTAENPEYL 1250
 DB 1195 VENPEYLTPOGGAAPQHPPPAPSPAFDNLVYWDQDPERGAPSTKGTPTAENPEYL 1254
 QY 1251 LDVPU 1255
 DB 1255 LDVPU 1259

RESULT 2
 Q9OX70 PRELIMINARY; PRT: 1209 AA.
 ID Q9OX70.
 AC Q9OX70.
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor.
 GN EGFR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RX MEDLINE=9025868; PubMed=2342466.
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
 RA Barch H.S.;
 RT "A truncated, secreted form of the epidermal growth factor receptor is
 RT encoded by an alternatively spliced transcript in normal rat tissue";
 RL Mol. Cell. Biol. 10:2973-2982(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Petch L.A.;
 RL Submitted (NOV-1991) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISHER; TISSUE=LIVER;
 RA Guttridge K., Dawson T.L., Barch H.S.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: M37394; AAF14008.1; -.
 DR HSSP: P13362; 1FGK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Dkinase.
 DR InterPro: IPR001245; Euk_Dkinase.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; kinase; 1.
 DR Pfam: PF01030; Recep_L domain; 2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SMO0261; Fu_3.
 DR SMART: SMO0219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00101; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
 SQ SEQUENCE 1209 AA; 134891 MW; 966EERFEC1B773 CRC64;

Query Match 44.5%; Score 3033; DB 11; Length 1209;
 Best Local Similarity 49.1%; Pred. No. 3; 7e-223;

Matches: 630; Conservative 166; Mismatches 362; Indels 124; Gaps 30;

QY 3 LALRCWGLLALLPFGA-ASTQVCTGTMKRLPASPETHLDMRLHYOGQVQVGNLE 61
 Db 15 LALCAAG-----GALBEKKVCGGTGNTNRLQGTDFEDHLSIQRMNNEEVLGNLE 66
 QY 62 LITLPTNNASISFQDIQEVQGYVLINNOVROVFLORLIVGTQLFEDNYALAVDNGD 121
 Db 67 ITVVQRNDISFQKTIOEVAGYVLIATLNTVERIPENIQIIRGNALYENTVALAVLSN-- 124
 QY 122 PLNNTPVTGASPGSLRELQRLSTLEILKGVLIOBNPOLCQODITLMDIFHKNQAL 181
 Db 125 -----YGTNTGTGRLPMRLQIILGAFRSNPLICMNETIQMRDIY-QDVFLSN 175
 QY 182 TLIDITNRS-RACHPCSPCKSGSRFCWSESSEDCQSLTRVACAGCA-RCKGPLPTDCHEQ 239
 Db 176 MSMDVQRHLTGCKPCDPCSPNGSCWGRGENCOKLTKIICAQCCSRCKGRSPDCCHNQ 235
 QY 240 CAAGCTGPKSDCLALHFNHSGICELHCBALVTNTDPEESMPNEEGRYTGASGVTAC 299
 Db 236 CAAGCTGPKSDCLALHFNHSGICELHCBALVTNTDPEESMPNEEGRYTGASGVTAC 295
 QY 300 PNYVLTSDVSGCTLVCPLEHNOEVTADGTQRCCKSPCARVCGYIGMOYIRANSKFTGI 359
 Db 296 PRNYVVTDHGSCVRACGPDYEV-EDGVSKCKKCGPCRYKONGIGEFK-DTISIVA 353
 QY 360 TELE-FAGCKKITGSLAFIPESFDGPAANTAPLOEQVQVETLEITGYLYISAMPDS 418
 Db 354 TNKHFKYCAISGDIHLPVAFKGSFRTPELDRLEIKTYVEIRIGFLLIQAAMPEN 413
 QY 419 LPDLVSFQNLQVIRGRILHNHGAISLTLOGISWLSRSLRELSGLALIHNTHLCFPH 478
 Db 414 WTDLHAFENLEIRGTQKHGQSLAVVGLNITSLSRLKEISDQDVIIISGRNLCYAN 473
 QY 479 TVPMDOLFERNPHOALHTANRPBECVGEGLACHOLCARHCKWGPPTQCVNCSQFLRQ 538
 Db 474 TTNMKLFGTPNOKTKIMNRAEKDCAIYNHVCNPLCSSEGCWGPPTQCVNCSQFLRQ 533
 QY 539 ECEEGCRVLQGLPREVYNARHCLPCHPECOQPNQSGTTCGFEADQVCAAHKDDPFCVA 598
 Db 534 ECVDKCNILEGEERREVENSECTQCHPECLPOTMNTTCGRGPDNTIKAAHVVDGHCCK 593
 QY 599 RCPSSGVKPLSYPIWKPFEDEGACQPCINCTHSCVDLDDKCPAECQASP-LTISVA 657
 Db 594 TCPSGIMGENNTL-VMKFADANNVCHLCHANCTYGCAGGLKGC-QQPEPKISISLNG 650
 QY 658 VGCILLVVLGVVFGI-LIKRQOKIRKTYMRLLDETLEVLTPSGAMENQAQMRITIK 716
 Db 651 IVGGILFTIV-VALGIGLFMRRLQVRRKTLRLLOERELVEPLTPSGEAPQAHRLIK 709
 QY 717 ETELKRVKYLGSAGFNNFTVSFWLRVFKVASASHELV-----LRENTSPKAKEILDEAY 772
 Db 710 ETEFKIKYLGSAFG-----TYKGLMIPGEKWKI PVALIKELRETSKAKEILDEAY 765
 QY 773 VVAGVSPYVSRLLIGLITSTVQVLTQMPYCGLLDHFENRGLGSDQLLWMCQIANG 832
 Db 766 VVASVNDPVCRLIGLITSTVQVLTQMPYCGLLDHFENRGLGSDQLLWMCQIANG 825
 QY 833 MSYLEDVRLVHRDLARVNLVSPNHVKITDGLARLLDIDETVHADGKXPIPMAL 892
 Db 826 KMYLEDRLVHRDLARVNLVSPNHVKITDGLARLLDIDETVHADGKXPIPMAL 885
 QY 893 SLRRRFTHQSDVSVYGVTVWELMTFGAKPYDGIAPRELPDLEKGERLPQPICITIDY 952
 Db 886 SLHRIYTHQSDVSVYGVTVWELMTFGSKPYDGIAPRELPDLEKGERLPQPICITIDY 945
 QY 953 MIMVKCMITDSECRPRELVSFMSMADPQRFVITQ-NEDLGPASPLDSTFYSLLED 1011
 Db 946 MIMVKCMITDSECRPRELVSFMSMADPQRFVITQ-NEDLGPASPLDSTFYSLLED 1005
 QY 1012 DMGGLVDAEEYLVPQGFCCPDPAFAGAGMHHRRSSSTRSGGDLTLGLPESEEAR 1071
 Db 1006 EDMEDVDADEYLTIPQGF-----NSPST-----S 1031

QY 1072 RSLPASEGAGSDYFDGLGGAAGLQSLPTHDPSPLOXYSEDPTVPLPSET--DGVA 1129
 Db 1032 RPLSLSSLSANSN-----SSTVACINRSGSCVXEDAFLOFYSDDPTSVLTEDNIDDFL- 1086
 QY 1130 PLTCSPOPEYVNOGDVAPRPPSPREGEPLPAPARPGATLERAKTISPGKGVYKVFYAFGG 1169
 Db 1087 -----PVEEYINQ-SVPRPAPASVQNPYYHQPLP-----APGRDLHYON--PHSN 1130
 QY 1190 AVENPEYL-TPQGAAPQHPHPAPFPAFDNLVYMDQ-----DP-----PERGA 1232
 Db 1131 AVSNPEYINTAQ-----PTLSSGFSBSALMIQKSHQMSLDNPDYQDFPKEAK 1181
 QY 1233 PESTFKGTPTAENPEYGLDVP 1254
 Db 1182 PNCIFKG-PTAENAYLRVAP 1202

RESULT 3
 Q9EP98 ID Q9EP98 PRELIMINARY; PRT; 1210 AA.
 AC Q9EP98;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Epidermal growth factor receptor isoform 1.
 GN EGF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Sinclair C.S., Pearse R.L., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maile N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Sinclair C.S., Pearse R.L., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Maile N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative Egr transcripts encoding truncated receptor
 RT isoforms."
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF275366; AAC28045.1; JOINED.
 DR EMBL; AF275365; AAC28045.1; JOINED.
 DR EMBL; AF275367; AAC24386.1; -.
 DR HSSP; P11362; 1FCF.
 DR MGI; MGI:95294; Egrf.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR000494; EGRF_L_domain.
 DR InterPro; IPR000719; Euk_Pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase; 1.
 DR SMART; SMO0261; Fu_5.
 DR SMART; SMO0220; S_TKc; 1.
 DR SMART; SMO0219; TyKc; 1.

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PSS0011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 KW ATP-binding; Receptor; Transferase.
 SC SEQUENCE 1210 AA; 134840 MW; 62CD021C9D32E18 CRC64;

Query Match 44.1%; Score 3004; DB 11; Length 1210;
 Best Local Similarity 48.6%; Pred. No. 6 2e-221;
 Matches 621; Conservative 168; Mismatches 368; Indels 120; Gaps 27;

QY 11 LLLALLPFGAA-STOYCTGTDMKRLRASPETHLMDRLHYOGCGVQGNLELYLPTN 68
 DB 14 LLLALLPFGAA-STOYCTGTDMKRLRASPETHLMDRLHYOGCGVQGNLELYLPTN 73
 QY 69 ASLSFLDIDIOGVGVYLAHNOYRQVPLQRLIVRGTLFEDNYALAVLDNGLPNTTP 128
 DB 74 YDLSFLDIDIOGVGVYLAHNOYRQVPLQRLIVRGTLFEDNYALAVLDNGLPNTTP 124
 QY 129 VTGASPGGLRELQRLSITELIKGVLIQRNPOLCYODTILMKDI---FKNNQALATLI 184
 DB 125 -YGTNRGLRELQRLSITELIKGVLIQRNPOLCYODTILMKDI---FKNNQALATLI 160
 QY 185 DTRSRACGECSPKCSRCWGSSEEDCCSLRTVACAGCA-RCKRPLPTDCHECCAG 243
 DB 181 -QSHPSGCPKCSRCWGSSEEDCCSLRTVACAGCA-RCKRPLPTDCHECCAG 239
 QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPGRYTFGASCTVACPVNY 303
 DB 240 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPGRYTFGASCTVACPVNY 299
 QY 304 LSTDVSGTCLVCPHLNQEVAEDGTQRCCKSPKRCVAVCYGLMGIKANSKFIETLE 363
 DB 300 VAVDHSQVACACPDYEV-EEGDIRCKCKDGPCKVCKGIGIEFFK-DTUSINATNK 357
 QY 364 -PAGCKKIFGSLAFLEPSFDDPASNTAPLOPELOVEFLEISITGVYISAMPDLEPL 422
 DB 358 HFRTCTAISGDLILPFAFGDSFTTRPDPRELELTKVKSITGELILOAPDMDTL 417
 QY 423 SVFQNLQVIRGILHNGAVSLTQGLISWGLRSRELDSGLALIHNTLFCFATVP 482
 DB 418 HAFENLEIRGTRKQGFSLAVVGINITSGLRLSKIEDGVIIISGRNLCYANTIM 477
 QY 483 DQLEFRPHQALLTARPREDECEGEGLAGHQLCARHGCMWPGTQCVNCSQIFRGECE 542
 DB 478 KRLFGPNQTKIMNNRKAERDCVAVHVCNPLCSSGCGWPEPRDCVSCQNSRGRECE 537
 QY 543 ECRVLQGLPREYVNAHCLPCHPECPONGSVTCFPEADQVACAHYKDPFCVARS 602
 DB 538 KNLLEGEPREREVENSECICHEPCLPQAMNITCTGRPNQICAHYIDGRHCVTCE 597
 QY 603 GVRPDLSTYPIKFPDEGACQPCINCHSCVDLDKGCAPAEORASPLTSIVSAVGL 662
 DB 598 GIMGENNTL-VMKYADANNVCHLCHANCYGCAGPGLQGEVWPSGPKPSIATGVGL 656
 QY 663 LVVVLGVVFGI-LIKRQCKIRKYTMRLLOETELVEPLTPSGAMPNOQMRILKETE 721
 DB 657 LFTIV-VAGIGIFMRNRHIVRKIRRLQRELEVEPLTPSGEAMQMRILKETE 715
 QY 722 KAVVLGSGAFGNFTVSWLAVPKVSASHLEV---LRENTSPKANKIILDEAVYAGV 777
 DB 716 KIRYLLGSGAFGNFTVSWLAVPKVSASHLEV---LRENTSPKANKIILDEAVYAGV 771
 QY 778 GSPYVSRLLIGITSTVQVLTQMPYGLLDHYENRGRIGSDILNMCQIAKGSYLE 837
 DB 772 DNHVACRLIGITSTVQVLTQMPYGLLDHYENRGRIGSDILNMCQIAKGSYLE 831
 QY 838 DVALVHRDLAARNVLYKSPNVKVTGGLARLLDIDETEHADGKVP-KWMALESILRR 897
 DB 832 DRLVHRDLAARNVLYKSPNVKVTGGLARLLDIDETEHADGKVP-KWMALESILRR 891
 QY 898 RFTHSDVMSYGVTVWELMTFGAKPDGIPAREIPLLEGEGLPPPICTIDVYIMTK 957

DB 892 IYTHSDVMSYGVTVWELMTFGSKPYDGPASDISISILEKGERLPPPICTIDVYIMTK 951
 QY 958 CMVLDSEKPRFRELVEFSMARDDPQRFVYIC-NEDLGKPSPLDSTFYRLDDDDMGD 1016
 DB 952 CMVLDSEKPRFRELVEFSMARDDPQRFVYIC-NEDLGKPSPLDSTFYRLDDDDMGD 1011
 QY 1017 LVDAEELVPOQGFPCPDPAAGAGVMVHRRSSSTRSGGDLTLGLEPSEEAARSPLA 1076
 DB 1012 VYDAEELVPOQGFPCPDPAAGAGVMVHRRSSSTRSGGDLTLGLEPSEEAARSPLA 1037
 QY 1077 PSEGASVFPDGLGMAKAGLQSLPHDPSPLQRYSEDPLVPSET--DGVAAPLTC 1134
 DB 1038 SLSATSN---NSVACINNGSCRVEDAFLLQRYSSDPFGAATVEDVIDDAFL----- 1087
 QY 1135 PPEVYVNPDPVAPPEPPEPPLPAAPAGATLERAKTLSPGKGVWDVPAFGAVENP 1194
 DB 1088 PVEVYVNPDPVAPPEPPEPPLPAAPAGATLERAKTLSPGKGVWDVPAFGAVENP 1136
 QY 1195 EYL-TPQGAAPQPPHPPAFSPAFDNLVYWDQ-----DP-----BERGAPPST 1237
 DB 1137 EYLNTAQ-----PCLSSGFSNPAWLOKSHQMSLDNPDYQDPFPEKTPNGIF 1187
 QY 1238 KGTFAENPEYLGADV 1254
 DB 1188 KG-PTAENAELRVAP 1203

RESULT 4

QYH40 PRELIMINARY; PRT; 1165 AA.
 ID QYH40
 AC QYH40;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Receptor tyrosine kinase proto-oncogene.
 GN XMRK.
 OS Xiphophorus xiphidium.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_Taxid=8086;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RX MEDLINE=98241172; PubMed=9582016;
 RA Dimitrijevic N., Winkler C., Welbrock C., Gomez A., Duschl J.,
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
 RT overexpression and mutational alterations."
 RL Oncogene 16:1681-1690(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIO PURIFICATION;
 RA Schartl M.,
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U53471; AAD10500.2; -.
 DR HSSP: F11362; IFGK.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR InterPro: IPR000494; EGFR_L_domain.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001899; Gram_pos_anchor.
 DR InterPro: IPR001245; Tyr_kinase.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; Kinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR PRINTS: PS00109; TYRKINASE.
 DR ProDom: PD000001; Euk_kinase; 1.
 DR SMART: SM00261; Fu; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE: PSS0011; PROTEIN KINASE DOM; 1.
 DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
 KW ATP-binding; Receptor; Transferase.
 SC SEQUENCE 1210 AA; 134840 MW; 62CD021C9D32E18 CRC64;

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 2.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 KW ATP-binding; Kinase; Tyrosine-protein kinase.
 SQ SEQUENCE 1165 AA; 129614 MW; 7F7EE38D877A74E CRC64;

Query Match 38.4%; Score 2617; DB 13; Length 1165;
 Best Local Similarity 44.5%; Pred. No. 2,6e-191;
 Matches 571; Conservative 164; Mismatches 385; Indels 164; Gaps 35;

1 MELAALCRWGLLALPPG-AAST---QVCTGDMKRLPASPEHRLDMLRYLGGCV 55
 4 LELLER-----LLLLLSIGRCSTDPKXVCGSTNQMM---LDNHYLKKMKMGSCNV 56
 QY 56 VQGNLEITLPTNASLFLDIOEVQGYVLIANQVPLQRLIRVGTQLFEDNYALA 115
 DB 57 VLENLEITTYQENQDLSFLQSIQEVGYVLIANNEVSTIPLVNLRLRQNLTYEGNFTLL 116
 QY 116 VLNGDPLNNTPTVGTASPGGLRELQRLSTELIKGGVLIQRMPLCYQDITLMQIFRK 175
 DB 117 VMSNYOK-NPSSP--DYVQVGLKQLSLNLTSLSGVAKVSHNPLCNVETIWMQIVDK 173
 QY 176 NNOLATLIDITNSRAHPSCSPCKSGRCWSESSDQSLTRTVCAQC-ARCKGFLPTD 234
 DB 174 TSNPTMMLIPHAFERQCKDCDPCGVNCSGAPRGHCQKFTKLCAEQCNRRCRGPKPID 233
 QY 235 CCEEQCAAGCTGRKSHDCLALPHNSGICELHCPALVTYNTDFESMRNPGRTYFGAS 294
 DB 234 CCNEHCAGGCTGRATDCLACRFNDGCTCKDCEPKKIYDIVSHVWNPNKXTYFGAA 293
 QY 295 CVTACFPYNLTSTVSGCTVCPHNOEVTAEQTCCKSKFCARCVGLGM---QYI 350
 DB 294 CVECECSNYVTE-GACVRCSCAGMLEVD-ENGRCKRPDQVCPKVCBGIGISLNTI 351
 QY 351 KANSKFTGTELEFAGCKIFGSLAPLPSFDCDPAISNTAPLOPEQLQVFTLEETGYL 410
 DB 352 ANVSTNIG---SFSNCTKINGDIILNRNFEEDGPHYKIGPMPEHMLMTYKEITGYL 407
 QY 411 YISAMPDLSLDFVFNLOVIRGRILHNGAYS-LTIQGLISMLGRSLRELQSLALH 469
 DB 408 VIMWPEMNTSLSVFQCLLEITIRKRTTPSRGFSFVVQVSHLQWLGSSLKESVAGVILK 467
 QY 470 HNTLFCFVITVPDQLFRNPHQALHTANRPEDECVGEGLAQHQLCARGHCMGPGTQCV 529
 DB 468 NTPQLRYASTINRRLFRSEDDQIEVDART-----ENQTCNNESSDEGCMGPGTMCV 520
 QY 530 NCQQLRQGECEVCEKVLQGLPREYVNAARHCLFCHPECCQONSVTGCFPEADQCVACH 589
 DB 521 SCHLVBRGGCVASCNLLQGEPRQAVDGRCVQCHOECIVQDTSITCTYGGPANCKSKAH 580
 QY 590 YKDPFCVACPSGVPRDLSYMPIMKFPDEGACQPCPICTHSCVDLDDKGPAPQGRAS 649
 DB 581 FQDGPQCIIRCPHMLGSDGTL-IMKADKMGCCQCHQCTGCGGPGISGGRGP-ITS 638
 QY 650 PLTSTISAVVGLLVVLLGVFGLIKRROQKIRKXTMRRLLOETELVEPLTPSGAMPNQ 709
 DB 639 HSSLAVGLVSGLLITVALLIVLLRRRIK-RKRTIRRLLOEKELVEPLTPSGAPNQ 697
 QY 710 AQNRILKETELKAVKVLGSGAFGNFTV--SPW-----LRPKXSAHLEXTRENTSP 761
 DB 698 AFRIKLETFKDRVLGSAFC---TYKGLMNDGGENIRLPVA---IKVLRATSP 749
 QY 762 KANKELIDEAYVAVGSPYVSHLIGLSTVQVLTQMLPYGCLLDHYENRGLSGSD 821
 DB 750 KVAQEVLDKAYVAVSDHPIVCRLLGLICLSAVQLVTLQMLPYGCLLDHYRQGERICGM 809
 QY 822 LNWCMQIAKMGVYLDVLYHNDLAARNTVKSPPHVKITDGLARLDDIETEVHAG 881
 DB 810 LNWMCVQIAKMGVYLDVLYHNDLAARNTVKSPPHVKITDGLARLDDIETEVHAG 869
 QY 882 GKVPYIKMALESILRRFTQSDVWSYGVTVWELMTFGAKPYDIPAREIPDLLEGERL 941
 DB 870 GKVPYIKMALESILRRFTQSDVWSYGVTVWELMTFGAKPYDIPAREIPDLLEGERL 929

QY 942 POPPICTIDVYMIWKKMMIDSECRPRERELVSSESMARDPQRFVNIQNEDLGAPSLD 1001
 DB 930 POPPICTIEVYMIILKCMWIDPSRPRERELVSSESMARDPQRFVNIQNEDLGAPSLD 986
 QY 1002 STFYSLLEDDEDDMGVJNAERYLVYQGFPCPDPAFGAGVNHHRHSSSTRSGGDLT 1061
 DB 987 RRLFRLLSSDD--DVVADLEYL-----NYKRI-NRQGS----- 1018
 QY 1062 GLEPSEERAPRSPPLAPSEGASDVFDGAGAAKGLQSLPTHPSPQRYSEDPV-PL 1120
 DB 1019 -----EPCIPRNGH-----FYRENSIALRYSIDPTQNAL 1047
 QY 1121 PSETGYVAPLTGCEQPEYVNPQVPRP-----PSPRE-----GPLP-AARPAQATL 1167
 DB 1048 EKDLIDGH-----EYVQPGSETSRLSDIYNPYEDLTDGMPGVSLSSQEAETNF 1097
 QY 1168 ERAKTLSPGKGVVADVAFGAVENPEYLTFGGGAQAPRPPAPFAPNLVYWDQDP 1227
 DB 1098 SRPEYLTNNQNSL---PLVSSGSDPDY---QAG-----YQAF-----L 1132
 QY 1228 PERGAPSTFKGTPRAENPEYLG 1251
 DB 1133 PQTGALTNGMFLPAENLEYLGL 1156

RESULT 5
 ID Q9MEF6 PRELIMINARY; PRT; 1137 AA.
 AC Q9MEF6;
 DT 01-NOV-1999 (Tremblrel. 12. Created)
 DT 01-NOV-1998 (Tremblrel. 12. Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21. Last annotation update)
 DE Receptor tyrosine kinase (fragment).
 GN ERBA4.
 OS Gallus gallus (Chicken).
 OC Archaeoptera; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_taxid=9031;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=HINDRAIN;
 RX MEDLINE=99263203; PubMed=10328884;
 RA Dixon M., Lumsden A., "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain." Mol. Cell. Neurosci. 13:237-258(1999).
 RL EMBL; AF121963; AAD31764.1; -.
 DR HSSP; P11362; 1FGK.
 DR InterPro: IPR000494; EGFR_L domain.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR002174; Euk_Pkinase.
 DR InterPro: IPR001368; Tyr_Kinase.
 DR InterPro: IPR001245; Tyr_Pkinase.
 DR InterPro: IPR004019; YAP motif.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01030; Recept_L_domain; 1.
 DR Pfam; PF02757; YAP; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_Pkinase.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TyrKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 KW PROSITE; PS00552; TNFR_NGFR_1; UNKNOWN_1.
 KW Kinase; Tyrosine-protein kinase.
 FT NON TER 1
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F0DC4F CRC64;

Query Match 37.9%; Score 2580.5; DB 13; Length 1137;

Best Local Similarity 45.5%; Pred. No. 1.6e-188; Matches 521; Conservative 177; Mismatches 358; Indels 89; Gaps 29;

```

QY 161 LCYODITLWKDIFHKNOALLTIDITRSRACHPCSMCKGSRGWSSSDCSLRTTV 220
DB 3 LCFADITHMODIVNPMASNFLLVPTNGSSGCGCHSKCTG-RGWGFTEHNCOTLTKTV 61
QY 221 AGGC-ARCKGFLPDCCHGEOCACTGPKHSDCLACHFNHSGICEHLCPALVTYNTDT 279
DB 62 AEQCDGRCYGPYVSDCCHREACGSCSPKOTDCACMNFDSGACTVQCQGTVPNPTTF 121
QY 280 ESMNPEGRYTFGASCTYACPYNYLSTDVGSCTLVCPHNOEYTAEDGTORCEKSPCA 339
DB 122 QLEHNHAKYTYGAFCKKCPHNFV-VDSSCVACAPSSKMEV-EENGIMCKRPTDIDCP 179
QY 340 RVYGLMNOYKAKSKFGLTELE-FAGCKKIFGSLFLPESFGDASNTAPLOPQLO 398
DB 180 KACDGTGSL-VSAQTVDSNIDKFNCTKINGNLFLVTGTHGDPYHITAIINPEKLN 238
QY 399 VFETLEITGYLYISAMPDLSPLDSVFQNTQVIRGRILHNGAVSLTQGLISWLGLRSL 458
DB 239 IFQVREITGYLNIQSPENMTDRVPSNLVTIGRATYSGSLILKQOQITSLQFQSL 288
QY 459 RELSGSLALHNNHLCGVHTVHPMDQLEFNNHQAULTANRPEDECVGEGLAGHQLCARG 518
DB 299 KQISAGNIYITDNLNLCYHTVNTSLFSTPSQKTVHRRKKAENCTADMCNCLCSD 358
QY 519 HGMGPGPQCVNCGQPLRGCEVCECVLQGLPREYVNAHCLPCHRECP-ONGSVTCF 577
DB 359 GCKGPGPDCLSCRFIRGRTICISCNLYGDFEFNANGSCMCDQCCKMDNMTTCY 418
QY 578 GREADQVACAHYDPPPCVARGSPGVKPLSYPIKFPDEBGAOCPPINCTHSCVDL 637
DB 419 GPGDHCTKCFHFGDNCVCEKCDGQGANFS--IFKYADEDECHPCPNCTOGGRGP 476
QY 638 DDGCG-----PAEGRASPLTSIVSAVV-GILLVYLVGVFGILLKRRQCKIRKY 685
DB 477 ASHCCTIYPTWRTOSTLQOHAR-TPL--IAAGVIGGLIYIMGLTFAYVRRKSIK-KKR 532
QY 686 TMRRLQETLEVEBLTBSGAMPNOQMRILKTELRLKRYKVLGSGAFENFTVSMFLRVP 745
DB 533 ALRRFL-ETELVEBLTBSGAPNOQMRILKTELKELKVKVLSGAFG---TVYKGIWVP 587
QY 746 -----KVSASHLEVLRENTSPKANKELIDEAYVWAGSPVYSLDICTSYVQVLTOL 800
DB 588 EGETVKLPVA-IKILNETTGPKAVFERDEBALIASHMDHPLVLLDVCUSPTIQLTOL 646
QY 801 MPYGLLDHVAENRGRGSDLLNMCQIAKMSYLEDVRLVHRDLAARVLYKSPNHVK 860
DB 647 MPHGLLDYVHEHNDNIGSOLLNMCQIAKMSYLEDVRLVHRDLAARVLYKSPNHVK 706
QY 861 ITDFGLARLDIDETEHADGKVPKIMMALESLRRRFTHQSDVMSYGVTVWELMTFGA 920
DB 707 ITDFGLARLDIDETEHADGKVPKIMMALESLRRRFTHQSDVMSYGVTVWELMTFGA 766
QY 921 KPYGIPAREIPDLLEKGERLPORPICIIVYIMINVCMMIDSECRPRELVESEFRMA 980
DB 767 KPYGIPAREIPDLLEKGERLPORPICIIVYIMINVCMMIDSECRPRELVESEFRMA 826
QY 981 RDPORFVYIQNEP-LGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCDPAPGA 1039
DB 827 RDPORFVYIQNEP-LGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCDPAPGA 885
QY 1040 GGMVHHRSSSTSGGGDLTLGLEPSESEAPRS--PLAP-SEGAGSDVFGDGMGAAK 1096
DB 886 RTIRDSNRNCFVYDGYAAEGV-PMYRARPGLIIBAPVAQATLEIFEDTCCNGTLR 944
QY 1097 GLGSLFTHDPSLPQYSEDPVPLPS-----ETDGVABLTCSPOPEYVNOBPAPOP 1149
DB 945 KQVATLAKEDSSSTQRYSDPTVFIIBERYIGELDEDDGYMTMRKPTDYLNVAENPFV 1004
QY 1150 PSPREGPLPA-RPARGATLERAKTSLSPKNGVVDVF-----AGGAVENPEYLTPO 1200

```

```

DB 1005 SRKKNQDQAVDNPEYHN-----APNGCPKADEYVNEPLYLNTFANTLENAEYL--- 1054
QY 1201 GGAAPQHPHPFSPAFNLYWQDDPERGA--PSTFKGPT-----AE 1244
DB 1055 -----KNMLPEAKAKAFNDPDYHNSLPPRSLQHPDYLQEXSTYFKQNGRIRPIYAE 1109
QY 1245 NPEYL 1249
DB 1110 NPEYL 1114

```

RESULT 6

```

ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE EtdB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OC NCBI_TaxId=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9317347; PubMed=1007531;
RA Gellner K., Brenner S.
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes".
RL Genome Res. 9:251-258(1999).
DR EMBL: AF056116; AAC34391.1; -.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000494; EGFR L domain.
DR InterPro: IPR000719; Euk kinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr kinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR ProDom: PD000001; Euk pkinase; 1.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00111; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647B9 CRC64;

```

Query Match

32.3%; Score 2204; DB 13; Length 1328;

Best Local Similarity 39.2%; Pred. No. 1.4e-159; Matches 508; Conservative 161; Mismatches 418; Indels 208; Gaps 36;

```

QY 9 WGLLALLP--GAATQ---VCTGDMKRLPASPETHLMDLHLHYOGCQVQGNTEL 62
DB 4 WRLLMCVASRLRAASSQGAVCPQTQGLSTSGQENYVNLNDRYGCEIIMNGLEI 63
QY 63 TYIPNTASLFLQIDIOEVGYLILANQVRQVPLQRLVVRGTQLPEDNYALAVDNGRP 122
DB 64 TQIESWDRSFLKTRIEVGYVLIANNHFOEIPLOQLRIRNSLYERFALSVPLN-- 120
QY 123 LNNVTVTASPGARELQRLSLTELKGVLIQRNPOLCYODITLWKDIFHKNOALAT 182
DB 121 ----YKDG--FSGNLQGLNMLTELDSGVQIINNKYIRYGPWYWRIL--RNDAPLE 173
QY 183 LIDTNRSRACHPCSPCKGSRGWSSSDCSLRTTVCAAGC-ARCKGFLPDCCHGEOCA 241
DB 174 IQNGERGVCH--KSC-GNYCWGPGKQCCQILITVCAPOCNDRCFGTSPDCHIECA 229
QY 242 AGGTGKSHSDCLACHFNHSGICELHCPALVYNTDTPESMNPBGRYTFGASCTYACPY 301
DB 230 AGCKGFLDTCFACRLFNDSGACVPOCPTLIYNKQTTOMETNPAAKYQYSGICVSGCPT 289

```

```

QY 302 NYLSTVDSCTLVCPHNDVETADGTOR-CEKSKPCARVCYGLMOYIKANSKFIGIT 360
DB 290 HFV-VDSGCVSVCPDPMKEV--ERGSORCELCISCPKVEGTAE-----QRTVDS 342
QY 361 ELE-FAGCKKIFGSLAFPEPSFDGDPASNTAPLQPEOLOVEETLEETITGLVYISAPD 419
DB 343 NIDSFNCTKIQSLHFLVTLGIDPKVVPUDAKKLEFRTVREITDILNQSFKL 402
QY 420 PDLVFTQNLQVIRGILHNGAVSLTQGLISWLAGLSRELGSALIHNTLCEFT 479
DB 403 NDLVSFSLTTIGRSLFKRFSIMWMRIPTLTSLGRSLREISDSGVYISQAHLCYHNT 462
QY 480 VPMDOLEFRNH-QALHTANRDECEVGEGLACHQICAGCHGCPPTQCVNCSQLEQG 538
DB 463 VMTTOLFRGSRVANSLSNRPAECVADGRVCDPLCSGCGPGPDCLCRNYSRG 522
QY 539 ECVEECRVLOGLEPREYVNAH-CLPGHECQPNQSVTCFEPADQCVACAYKDPFCV 597
DB 523 TVAGCHFNNGIIPREFAGLNGVAVACHPECKPOTGASCTGPGADCMACTFRDQPYCM 582
QY 598 ARCPGVKPDLSMPYWKPFDEEGACQPCPTNCTHSCVDLDDKGCPRASPLTSIVA 657
DB 583 SSCPAGVN-DEGKGLIFKFPNREHCEPCHQCTGCGSGGLNDC--LEAARLTSSGQ 638
QY 658 VVGILLVVLGVVF-----GILIKRQOKIKRYTMRLLQETELVPLTSGAMPNA 710
DB 639 ITGIALGVPAGLIFCLVFLGLMHLRGLAIRKRAMRRLYESGESFEPLGP-GEKXTYV 697
QY 711 QMELKETEELKXKVLGSGAFGNNFTVS--FWL-----RVKVSASHLEVLRENTSPK 762
DB 698 HARLKPSDLRKIKPLGSGVFG---TVSKGFWIPEGETKIPVA---IKTIQDSGQ 749
QY 763 ANEILDEAVMAGVSPYVSRLLGICLTSVQVLYQLMYPGLLDHVENGRGLSQDL 822
DB 750 TTEETIDHLISKSLHPYVRLGLICPGCLQVLQVLSHSGSLHIIHQKTSIDPRL 809
QY 823 LNMOMIAGMSYLEDVRLVHRDLAARNLVKS PNHVKITDFGLAELDIDETEHADG 882
DB 810 LNMVQIAGMYLLEHRVVKHKLARNILKNDYQVQISDYVADLLVPDCKYVYSFT 869
QY 883 KVIKMMALLESIRRFTHOSDVSQVYWEIMTGMKAPYDGIIPAREIPDLLEKGERLP 942
DB 870 KPIKMMALLESIRRFTHOSDVSQVYWEIMTGMKAPYDGIIPAREIPDLLEKGERLP 929
QY 943 QPRICTIDYVMIWVKCMITSECRPRELVSFSPMARDPQRFVYIQVEDGAPSLDS 1002
DB 930 QPAICTIDYVMIWVKCMITSECRPRELVSFSPMARDPQRFVYIQVEDGAPSLDS 980
QY 1003 TFYRSLEDDDGDLVDAEYLVPOQGFCCPDPAAGGMVHHRRSSSTRSGGDLTG 1062
DB 981 -----EDSGMGEFL-----RGRSER---GLLEAD 1001
QY 1063 LEPSEEPARSPILAPSEAGSDVFDGDLGNG--AAKGIQSLPTHPSLQ----- 1110
DB 1002 LEPDEEB-----GLGDRFAIPLSLOPSPSSWSTSSQINSYWMVMTQ 1040
QY 1111 -RYSDEPVLPESETDYVAFPLCSPOP-EYVNO-----PDVFPQPSPRE 1154
DB 1041 LRYD-----FAVSGGHIGYILPMSPSPYDITRQLWYQSRSLSSVTLDRSAFRSSSE 1094
QY 1155 GFL--PAARPAKTLERAKTLSGKRGVVKDVAFEGAVENPEYLTLPQGAAPQPHPPA 1212
DB 1095 AELCEDGACAGIFRVR-----FESERGN-----FOGG----- 1122
QY 1213 FSPAFNDLYWDDPPERGAPSTFKGTPTAENPE 1247
DB 1123 -----QQRKLTSTASSPESFKTMADEDE 1146

```

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4.
RA Lycett G.D.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor."
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ31655; CAC35008.1;
DR HSRP; P13362; 1FGK.
DR InterPro; IPR000345; Cytochrome_c.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINSE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; Fu; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN.4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON TER.
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
Query Match 28.2%; Score 1923.5; DB 5; Length 1433;
Best Local Similarity 31.6%; Pred. No. 4.6e-138;
Matches 459; Conservative 197; Mismatches 400; Indels 397; Gaps 40;
QY 26 CTGDMKRLPASPEHLDMLRHLYQCGQVYQGNLELTPTNASLSLFDIQEYQVYL 85
DB 1 CIGTNGKSVAPNREYHYKNLDRYTNCTYVDGNLEITWIONITDNLFIQHIREVTGYVL 60
QY 86 IAHNOVRQVPLORIRIVAGTQLF-----EDYVALAVLDNGPPLNNTTFTVGTASPEGAREL 140
DB 61 ISLYLPOVILPRLOIIRGRTTFKLNKKEAYGLFV-----SFSHNTL 104
QY 141 QLRSLTEILKSGVLIQRNPOLCYQDTILMKDI-FHKNNQALTLIDTNRSACHGSPMC 199
DB 105 ELPALRDLILGSGVGFENNYNLCMKMSIMWEILLAPQSMOYTNFSSPEKVCPCPHSC 164
QY 200 KGSRCWSSSSDQSLFTVACAGCA--RCGSLPLTDCQCHQCAACGPRGSHCLALH 257
DB 165 EVG-CWGGGAINCQFSLKNCSPQSGRCGPRFECCHLFCAGGCTGPRQSOCLACKN 223
QY 258 FNHSGICELHCPALVYNTDTFESMPNEBGRYTGASCVTAPYNYLSTVDSGCTVCP 317
DB 224 FYDDGVCKQCECPMQIYVPTNYFNEPNDGKAYAGATCVRCP-EHLKMGACVRCQPK 282
QY 318 HNOETADGTQRCCEKSKPCARVCYGLMOYIRANSKFJITELERFAGCKKIFGSLAFL 377
DB 283 GKMPQNSB-----CVPCGKGVCKTCPEGGI-----VHSDNTG-----NXYDCTIIGSLEIL 329
QY 378 PESFDGDPASNT-----APLQPEOLOVEETLEETITGLVYISAPDPLSVFQNLQ 429
DB 330 DQSFDDGQOVYVNTNPSFGRIYIKIDPDLLEVSTVYKEITGFINIOAHHPNFTLLVYFNLE 389

```

RESULT 7
 Q9BIH9 PRELIMINARY; PRT; 1433 AA.
 AC Q9BIH9;

```

QY 430 VIRGRILHNGAY-SLTLOGIGISWLGRLSRLSGSLALIHNNHLCFVHTVPMQDFEN 488
DB 390 VVGGRQJENKENVASYIVKTSLSKSLKSLKRVNSSVILESDLCFEDDIDWSITKRS 449
QY 489 PHQALHTANREDECEVGEGLACHOLCARGHCWGPPTOCVNCSCPLRQGECEVBEGRVLQ 548
DB 450 SDHEVAVQKRNNTTECHBEGMECSBQCSKAGCKGKPEQCLEKNKVKYKXCLDSCK--- 506
QY 549 GLFREY-VNARHCLPCHRECPQONSVYTCGPADOCVCAHAKDPFCYVACP----- 601
DB 507 SLPRLYSVDSKTCGDCHQCKD-----FCYGPNEHDCGSMVYKDRFCVACBPTTKHAM 561
QY 602 -----SGVKPDSLYMFKFPD----- 618
DB 562 NGTCINCKTCVCGRGRDTIADPGCISCDKALISDAKIERCLMDSCPDGYSYDVL 621
QY 619 -EEG----- 621
DB 622 QEEGPKLQLSGKAVCRKCHRCCKCTGYGFHEQFCOECTGYKKGEQCEDECPDYANEE 681
QY 622 -ACQPCINCT-----HSCVDL-----DD-----KCGPAEQ----- 646
DB 682 TRICLPCHQCRCHGIGDDHNECRNKLFEQDPYDNATFTCVSCNPASHPKYKPEQEA 741
QY 647 -----RASPLTISAVAVGILLVVLGVVFGI---LKKRQCKIRKTYM 687
DB 742 GKICPYCASDMSQGLRLEPOTQVKIYVSMALILLCVFGAFAVLFSHKCKKDAVVM 801
QY 688 RLLQETELVEPLTPSGAMPNQAOMRLKETELKXVYVGSFGFNNFVFSFVL----- 742
DB 802 TMLAGCEDESEPLRPSVGNPLTLRIKEAERIRGVLQMGAFG--RFFKGVWMEGSS 859
QY 743 -RVPKVSASHLEVLRENTSPKANKELIDEAVYVAGVSPVYSRLGICLTSTVQVLTOLM 801
DB 860 VKIPVA-----IKVLMEMSGSESSKEPLEEAVYVASHVPHLKLILAVCMSCQMLLTQLM 915
QY 802 PYGCLLDHRENRGRSLSDLLWCMQIAKMSLLEVLRYLVRDLAARNVLYSPNHVKI 861
DB 916 PLGCLLDYVNNKDKISSKALLNWSQIARGMAVLEERRVLRVHADLAARNVLYVTPCCKI 975
QY 862 TDFGLARLLDIDETEYADGKGVYIKMALESILRRFTQSDVWSYGVTVMLMTFGAK 921
DB 976 TVEGLAKLLDPSERYAAGKXPIKWLALBECIRHRYFTSKSDVMAGIITWELLTYGAR 1035
QY 922 PYDGIPIAREIPDLLEKEERLPQPICTIDVYMIWKCMTDSECRPFRELVSEFSMAR 961
DB 1036 PYENVPAKDVPELIEIGHKLPQDPICSLDVYCIILSCVLDADARPTFKLAETFAEKAR 1095
QY 982 DPORFVYIQNEDLGPASPLSTFYRSLLEDDMDGLY----- 1018
DB 1096 DPGRYLMI-----PGDKFMRLPSTYNODEKDLIRTLAPVMAAAAAAAGASNVLD 1146
QY 1019 -----DAEYLVBOQGFCCPDPAFGAGVHNRHSSSTRSGGDLTLGLPSEEBAPR 1072
DB 1147 VPSTIAETDEYLCQKTRPSIMLPQPSA-----VEPS-DEMPK 1182
QY 1073 S-----PLAF-----SEAGSDYFPQDGLMGAKLQSLPTHDSPLQRYSEDPVLPSE 1123
DB 1183 SLRYCKDPLKPDDETDGHEKVV-----GVGGR-----LNLPLD 1216
QY 1124 TDGYVAPLTCSPQPEYVNPDPVRPQPSPREGULPAARPAQATLERAKTSLPGKNGVVKD 1183
DB 1217 EDDYLMPTCQSQ-----NQS-----TPG-----YMD 1237
QY 1184 VFARFGAVENPEYL-----TPQGAAPQHPPPAFSAFNNLYVMQDPPERCAPR 1234
DB 1238 LIGVAVSDNPEYLMGSTQALAGLQSGMG--PHTPP-----PP 1274
QY 1235 STFKGPTAENPE 1247
DB 1275 NTPNGMPTHQHSQ 1267

```

```

RESULT 8
Q9UK79 ID Q9UK79 PRELIMINARY; PRT: 419 AA.
AC Q9UK79,
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RL autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; F6C1BE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 7,7e-135;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELALCRKGLLALLPFGAASVQCTGDMKRLPASPETHLMDLRHLQCCQVVOGVL 60
DB 1 MELALCRKGLLALLPFGAASVQCTGDMKRLPASPETHLMDLRHLQCCQVVOGVL 60
QY 61 ELTYLPTNASLSFLDIOEVQGVYIAHNOVQVPLQRLRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVYIAHNOVQVPLQRLRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTIIMKDLFHKNOQLA 180
DB 121 DPLNNTTPTGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTIIMKDLFHKNOQLA 180
QY 181 LTLIDNRRBACHPCSPMKGSRGSESEDCSLTRFYCAGGACARCKGPLETDCHEQC 240
DB 181 LTLIDNRRBACHPCSPMKGSRGSESEDCSLTRFYCAGGACARCKGPLETDCHEQC 240
QY 241 AAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTFESNPNREGRYTFGASCVTACP 300
DB 241 AAGCTGPKASDCLACLFHNSGICELHCPALVTYNTDTFESNPNREGRYTFGASCVTACP 300
QY 301 YNYLSTDVSCCTLVGCLHNOEYTAEDGQRCCKSPCARVYXGL 345
DB 301 YNYLSTDVSCCTLVGCLHNOEYTAEDGQRCCKSPCARVYXGL 345

RESULT 9
Q9R2X1 ID Q9R2X1 PRELIMINARY; PRT: 367 AA.
AC Q9R2X1,
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;

```

BACKFILE DOCUMENT INDEX SHEET

REV 7/02/03

A DOCPHOENIX



APPL PARTS

IMIS
Internal Misc. Paper
LET
Misc. Incoming Letter

371P
PCT Papers in a 371 Application

A...
Amendment Including Elections

ABST
Abstract

ADS
Application Data Sheet

AF/D
Affidavit or Exhibit Received

APPENDIX
Appendix

ARTIFACT
Artifact

BIB
Bib Data Sheet

CLM
Claim

COMPUTER
Computer Program Listing

CRFL
All CRF Papers for Backfile

DIST
Terminal Disclaimer Filed

DRW
Drawings

FOR
Foreign Reference

FRPR
Foreign Priority Papers

IDS
IDS Including 1449

SRNT
Examiner Search Notes

CLMPTO
PTO Prepared Complete Claim Set

Internal

ECBOX
Evidence Copy Box Identification

WCML
Claim Worksheet

WFEE
Fee Worksheet

OUTGOING

CTMS
Misc. Office Action

1449
Signed 1449

892
Abandonment

ABN
Board of Appeals Decision

APDEC
Examiner Answer

CTAV
Count Advisory Action

CTEQ
Count Ex parte Quayle

CTFR
Count Final Rejection

CTNF
Count Non-Final

CTRS
Count Restriction

EXIN
Examiner Interview

M903
DO/EO Acceptance

M905
DO/EO Missing Requirement

NFDR
Formal Drawing Required

NOA
Notice of Allowance

PETDEC
Petition Decision

INCOMING

AP.B
Appeal Brief

C.AD
Change of Address

N/AP
Notice of Appeal

PA..
Change in Power of Attorney

REM
Applicant Remarks in Amendment

XT/
Extension of Time filed separate

FWCLM
File Wrapper Claim

IFFW
File Wrapper Issue Information

SRFW
File Wrapper Search Info

File Wrapper